

Europäisches Patentamt

European Patent Office

Office européen des brevets



(1) Publication number: 0 541 335 A1

/NFCOTK 15/00F12812

(12)

EUROPEAN PATENT APPLICATION

INF CO7K 15/00 F12B 20 AZA

21) Application number: 92310067.1

(5) Int. CI.5: C12N 15/62, C12N 5/10

(2) Date of filing: 04.11.92

INF A61439/02

(30) Priority: 08.11.91 US 792507

Date of publication of application: 12.05.93 Bulletin 93/19

(84) Designated Contracting States : CH DE FR GB IT LI NL

(1) Applicant: MERCK & CO. INC. 126, East Lincoln Avenue P.O. Box 2000 Rahway New Jersey 07065-0900 (US)

7

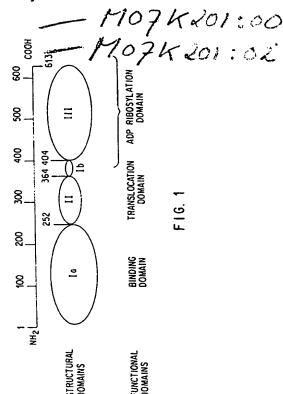
172 Inventor: Donnelly, John J.
1505 Brierwood Road
Havertown, PA 19083 (US)
Inventor: Liu, Margaret A.
4 Cushman Road
Rosemont, PA 19010 (US)
Inventor: Friedman, Arthur
121 Froghollow Road
Churchville, PA 18966 (US)

Inventor: Marshall, Mark S. INF HOIK 39/12 1519 Spruce Court Carmel, Indiana 46032 (US) WFC12N 5/08
Inventor: Hawe, Linda A WFC12N 5/08 2610 Skippack Pike Norristown, PA 19403 (US) Inventor: Montgomery, Donna L. 9 Hickory Lane Chalfont, PA 18914 (US) Inventor: Oliff, Allen A. 1412 Florence Drive Gwynedd Valley, PA 19437 (US) Inventor: Shi, Xiao-Ping 536 Winthrop Road Collegeville, PA 19426 (US) Inventor: Ulmer, Jeffrey 128 Dolly Circle Chalfont, PA 18914 (US)

(74) Representative: Thompson, John Dr. et al Merck & Co., Inc. European Patent Department Terlings Park Eastwick Road Harlow, Essex CM20 2QR (GB)

Recombinant DNA sequences and plasmids for cellular immunity vaccines from bacterial toxinantigen conjugates.

Recombinant DNA sequences coding for hybrid proteins having two primary components. The first component is a modified bacterial toxin that has translocating ability, while the second component is a polypeptide or protein that is exogenous to an antigen-presenting cell. The hybrid has the ability to be internalized by an antigen-presenting cell, where the hybrid is subsequently processed and an antigenic segment of the hybrid presented on the surface of the antigen-presenting cell, where the segment elicits an immune response by cytotoxic T lymphocytes.



### BACKGROUND OF THE INVENTION

5

10

15

20

25

30

35

45

50

The numerous substances and organisms that threaten the existence of animals having immune systems are either present in extracellular body fluids, such as toxins or bacteria, or else they are harbored within the animal's own cells, such as viruses, certain parasites and oncogene products. This distinction is important to thymusderived lymphocytes, also known as T cells, which are an important component of vertebrate immune systems. T cells have evolved parallel systems for recognizing intracellular and extracellular antigens. In both systems, antigens are recognized only when they are bound to molecules of the major histocompatability complex (MHC).

The MHC encodes two types of cell surface molecules that act as receptors for protein antigens. Class I MHC molecules consist of a highly polymorphic integral membrane glycoprotein alpha chain that is noncovalently bound to a beta<sub>2</sub> microglobulin. Class II MHC molecules consist of two noncovalently bound, highly polymorphic, integral membrane glycoproteins. Class I MHC molecules have a groove at the top surface formed by the two amino-terminal domains. The groove holds an antigen. As with other cell surface proteins, during cellular processing in the cytosol, MHC molecules are inserted into the endo-plasmic reticulum (ER) and, following chain assembly, are transported to the plasma membrane of the cell via the Golgi complex and post-Golgi complex vesicles.

The recognition of Class I vs. Class II molecules as antigen-presenting sites in general divides T cells into two classes, respectively termed cytotoxic T cells ( $T_c$ ) and helper T cells ( $T_H$ ).  $T_C$  cells directly lyse cells that are infected with viruses or certain parasites and also will secrete cytokines such as gamma-interferon in order to eradicate intracellular pathogens and tumors.

Virtually all cell types can serve as antigen-presenting cells for  $T_{\rm C}$  cells as long as they express MHC Class I molecules. In general,  $T_{\rm C}$  cells require antigen-presenting cells that are actively biosynthesizing antigen. During processing, the antigen is bound to a nascent Class I molecule in the ER and transported to the plasma membrane via the Golgi complex and post-Golgi complex vesicles. At the plasma membrane, the processed antigen sits in the groove of the MHC Class I molecule, where the processed antigen is available for binding to cell surface receptors of  $T_{\rm C}$  cells. Activation of  $T_{\rm C}$  cells requires interaction between multiple  $T_{\rm C}$  cell surface molecules and their respective ligands on antigen-presenting cells. Once activation has taken place, the lysing and cytokine secretion activity described above can begin.

Antigen processing is the structural modification and trafficking, within the proper subcellular compartments, of protein antigens that enable the determinants recognized by  $T_{\rm C}$  cells to interact with MHC molecules. As noted above, most, and possibly all, somatic cells expressing MHC Class I molecules constitutively process antigens and transport determinants to the cell surface for  $T_{\rm C}$  cell recognition. Antigen processing is thus required for the presentation of intact, folded proteins to  $T_{\rm C}$  cells. Commonly, antigen processing entails the generation of short peptides by cellular proteases, although some intact proteins productively associate with MHC molecules, indicating that proteolysis is not necessarily a component of antigen processing.

Two distinct pathways are used by cells to process antigens. The endosomal pathway is so named because it is accessed through the endosomal compartment. Determinants produced by this pathway usually associate with Class II MHC molecules. The other pathway is the cytosolic pathway. The cytosolic pathway is so named because it can be accessed from the cytosol of the cell by the synthesis of proteins within the cell, or by penetration of plasma or endosomal membranes by extracellular proteins. Such penetration may occur naturally through the fusion of the cell's membrane with a virus, or artificially by osmotic lysis of antigencontaining pinosomes. Determinants produced by cytosolic processing typically associate with Class I MHC molecules. The cytosolic pathway is able to process many different types of foreign proteins for presentation to T<sub>C</sub> cells.

Class I MHC molecules associate with antigens in a compartment of the ER. In this regard, it is important to note that the compound Brefeldin A acts by interfering with the normal vesicular traffic between the ER and the Golgi apparatus, and thus also has the effect of blocking the presentation of cytosolically processed antigen on the surface of what would otherwise be an antigen-presenting cell.

It can be seen from the above discussion that, in order to generate response by a cytotoxic T cell, it is generally necessary either to cause the target cell, which has been chosen as an antigen-presenting cell, to endogenously synthesize the protein antigen of interest, or to deliver exogenous protein antigen of interest directly into the cytosolic antigen processing pathway of the target cell. If the latter could be accomplished, a vaccine could be produced which would elicit cytotoxic T cells capable of killing virally or parasitically infected cells or tumor cells, thereby having particular usefulness for preventing three clinical types of diseases.

First, such vaccines could prevent infections caused by viruses such as papilloma or herpes virus which do not undergo a blood-borne phase of infection. This would be especially true in the case of human papilloma virus E7 protein, which is continuously cellularly expressed in the transformed phenotype, and would thus be

particularly well suited to attack by sensitized cytotoxic T lymphocytes.

Secondly, there are those infections caused by viruses such as influenza or human immunodeficiency virus (HIV) or parasites whose outer proteins may have high antigenic variability making it difficult to design a vaccine capable of eliciting protective titers of high affinity antibodies with broad specificity. Certain viral internal proteins have less antigenic variation, and peptides derived from such proteins when associated with Class I MHC molecules, would render infected cells susceptible to lysis by sensitized cytotoxic T lymphocytes.

Thirdly, tumors and virally transformed cells express neoantigens that may be presented on Class I MHC molecules, thus rendering these cells suitable targets for cytotoxic T lymphocyte lysis.

Current vaccines generally focus on generating humoral (that is, antibody) responses of the immune system, rather than the cellular immune responses discussed above. Those that do generate cellular immune responses use attenuated live viruses which replicate intracellularly, introducing their constituents into an infected cell's antigen processing pathway as a result of being synthesized within the cell thereby being available for the appropriate protein processing pathway. Thus, there is a need for a non-replicating vaccine that will sensitize cytotoxic T lymphocytes to produce a cellular immune response with a significantly greater margin of safety.

The present invention meets this need by capitalizing on the ability of certain bacterial exotoxins to be internalized into cells through endocytosis via receptors on the cell surface and then translocate out of the resultant endosomes into the cellular compartment in which endogenous proteins are processed for presentation. These exotoxins have been hybridized with polypeptide or protein antigens, which are carried into the cytoplasm and are processed to peptides capable of association with Class I MHC molecules via the physiologic processes discussed above. Once associated with a Class I MHC molecule and presented on the surface of the antigen-presenting cell, they can sensitize cytotoxic T lymphocytes against other infected cells synthesizing the same polypeptide or protein. By virtue of these actions, the invention presents vaccines which can be effective in prophylaxis against viruses, parasites and malignancies.

It is an additional object of the present invention to produce hybrid proteins of certain bacterial exotoxins having translocation domains, hybridized with polypeptides or proteins selected for their antigenic activity, which hybrids will be useful as probes for studying the intracellular processing and subsequent presentation of endogenously synthesized cytoplasmic proteins.

#### BRIEF DESCRIPTION OF THE DRAWINGS

5

10

15

20

25

30

35

40

45

50

55

Figure 1 shows the structural domains of <u>Pseudomonas</u> exotoxin, along with the numbers of the amino acid residues that define the known limits of the structural domains. Amino acid residues are numbered as defined in Gray, et al, PNAS USA <u>81</u> = 2645-2649(1984).

45.0

Figure 2 is a restriction map for plasmid pVC45-DF+T.

Figure 3 is a restriction map for plasmid pBluescript II SK.

Figure 4 is a restriction map for plasmid pBR322.

Figure 5 is a graph showing the results of using hybrid construct PEMa in immunologically sensitizing U-2 OS cells, a human cell line.

Figure 6 shows that a hybrid protein made of the binding and translocating domains of <u>Pseudomonas</u> exotoxin and a peptide epitope of influenza A matrix protein can competitively prevent the intact <u>Pseudomonas</u> exotoxin from binding to and killing target cells.

### SUMMARY OF THE INVENTION

The invention is the recombinant DNA sequences coding for a hybrid protein of two species, the first species being a modified bacterial toxin that has a translocating domain. The second species is a polypeptide or protein. The polypeptide or protein is exogenous to an antigen-presenting cell of interest. The hybrid of the bacterial toxin and the exogenous polypeptide or protein are constructed in such a way as to be capable of eliciting an immune response by cytotoxic T lymphocytes. Also included are suitable plasmids and methods of using the recombinant sequences to obtain the hybrid proteins of interest.

A preferred bacterial toxin is a modified <u>Pseudomonas</u> exotoxin. <u>Pseudomonas</u> exotoxin is known to consist of four structural domains, namely Ia, II, Ib and III. This is shown at Figure 1, along with the numbers of the amino acid residues that define the known limits of the structural domains. More preferably, the <u>Pseudomonas</u> exotoxin is modified by deletion of structural domain III, that is the ADP-ribosylating structural domain, although alternatively domain III need not be entirely deleted, but may rather be sufficiently altered in its amino acid sequence so as to render it enzymatically nonfunctional as an ADP-ribosylating enzyme. Most preferably, the modified bacterial toxin has only a cellular recognition domain and a translocating domain, (with or without

the 5 C-terminal amino acids of Domain III added to the C-terminus of the polypeptide or protein antigen), or even just the translocating domain with or without targeting ligand. In the case of <u>Pseudomonas</u> exotoxin, the cellular recognition domain and translocating domain are known to exist within structural domains Ia, II and Ib. Also most preferably, modified <u>Pseudomonas</u> exotoxins are arranged on the amino-terminal side of the hybrid, while the exogenous polypeptide or protein is arranged on the carboxyl-terminal side of the hybrid.

The exogenous polypeptide or protein, which is exogenous to an antigen-presenting cell of interest, is preferably a polypeptide or protein of viral origin. More preferably, the viral polypeptide is a viral protein fragment, and most preferably is taken from the group comprising the matrix protein of influenza A virus; residues 57 to 68 of the matrix protein of influenza A virus (the matix epitope known to bind MHC HLA-A2); the nucleoprotein of influenza A virus; or the GAG protein of human immunodeficiency virus-1.

Functionally, the hybrid is capable of eliciting an immune response by cytotoxic T lymphocytes, by virtue of being at least partially presented on an antigen-presenting cell surface. More specifically, the hybrid functionally is capable of being internalized by an antigen-presenting cell and further capable of being processed, via the endogenous protein processing pathway, on its way to at least partial presentation on the surface of the antigen-presenting cell.

The hybrid proteins preferably will use polypeptide or protein-antigens for use as a vaccine, and most preferably will use viral antigens. Most preferably, these viral antigens will be conserved viral proteins. The hybrids will be incorporated in an amount sufficient to elicit an immune response by cytotoxic T lymphocytes into vaccines further comprising pharmaceutically acceptable carriers. The vaccines will be sufficient to immunize a host against the diseases influenza, acquired immunodeficiency syndrome, human papilloma virus, cytomegalovirus, Epstein-Barr virus, Rota virus, and respiratory syncytial virus, tumors and parasites.

The present invention further relates to recombinant DNA segments containing nucleotide sequences coding for the fused proteins described above, as well as plasmids and transformants harboring such recombinant DNA segments, as well as methods of producing the hybrid proteins using such recombinant DNA segments and methods of administration of the hybrid proteins as vaccines to hosts.

# DETAILED DESCRIPTION OF THE INVENTION

5

10

15

20

25

30

35

45

50

55

The term "translocating domain" shall mean a sequence of amino acid residues sufficient to confer on a polypeptide or protein the ability to translocate across a cell membrane into a cellular compartment for processing endogenous proteins.

The term "exogenous to an antigen-presenting cell" shall mean polypeptides that are not encoded by the unmutated genome of a given antigen-presenting cell.

The term "antigen-presenting cell" shall refer to a variety of cell types which carry antigen in a form that can stimulate cytotoxic T lymphocytes to an immunologic response.

The term "immune response" shall mean those cytotoxic processes of cell lysis and cytokine release engaged in by cytotoxic T lymphocytes that have been stimulated by antigen presented by an antigen-presenting cell. This term shall also include the ability of a host's cytotoxic T lymphocytes to retain their cytotoxic response to subsequent exposure to the same antigen that will lead to more rapid elimination of the antigen than in a non-immune state.

The term "presented on an antigen-presenting cell surface" shall mean that process by which an antigen is seated within a ligand site of a major histocompatability complex Class I protein on the surface of an antigen-presenting cell.

The term "being internalized by an antigen-presenting cell" shall mean the process of endocytosis resulting in endosome formation.

The term "cellular recognition domain" shall mean a sequence of amino acid residues in a polypeptide sufficient to confer on that polypeptide the ability to recognize a receptor site on the surface of a target cell.

The term "ADP ribosylating domain" shall mean a sequence of amino acids sufficient to confer on a polypeptide the ability to modify elongation. factor II within a cell, and thereby severly impair the viability of the cell or kill it.

The term "vaccine" shall mean a pharmaceutically acceptable suspension of a given therapeutic entity administered for the prevention, amelioration or treatment of infectious diseases.

The term "conserved viral protein" shall mean those viral proteins that do not vary from strain to strain of a given species of virus, or to those viral proteins that are generally unlikely to undergo mutation as a function of time in a given strain.

The term "arranged on the amino terminal side of said hybrid" shall mean that a peptide sequence has been inserted at any point between the amino terminus of a hybrid and the hybrid's middle amino acid residue. The term "arranged on the carboxy terminal side of said hybrid" shall mean that a peptide sequence has

5

10

15

20

25

30

35

40

45

50

55

been inserted at any point between the carboxy terminus of a hybrid and the hybrid's middle amino acid residue.

The term "transformant" shall mean an independent, self-replicating DNA molecule, and shall include plasmids.

The hybrid proteins of the present invention are fusion protein constructs of a bacterial toxin having a translocating domain fused to a polypeptide or protein that has been selected for its antigenicity for a given disease, as well as for being exogenous to a targeted antigen-presenting cell. A preferred bacterial toxin is the <u>Pseudomonas</u> exotoxin. This exotoxin is known to comprise four structural domains, as shown in Figure 1. These domains are designated Ia, II, Ib and III. Structural domain Ia is known to be necessary for binding of the exotoxin to a receptor site on the surface of a target cell. Structural <u>domain II</u> is known to be necessary for translocation of the exotoxin across an internal membrane the targeted cell. Part of structural III are known to be an ADP ribosylating enzyme that bind to the protein Elongation Factor 2, which generally results in the death of the target cell.

In a preferred embodiment of the present invention, structural domain III (or all domain III except for the C-terminal amino acids) has been deleted from the Pseudomonas exotoxin molecule, and has been replaced with one of several polypeptides or proteins chosen for their ability to act as antigens and therefore be useful as vaccines. The antigens used for vaccines include antigens of viruses whose hosts are higher vertebrates, such as antigen of influenza A virus, human immunodeficiency virus-1, human papilloma virus, cytomegalovirus, Epstein-Barr virus, Rota virus, and respiratory syncytial virus. Other viruses include herpes viruses such as herpes simplex virus, varicella-zoster virus, adult T cell leukemia virus, hepatitis B virus, hepatitis A virus, parvoviruses, papovaviruses, adenoviruses, pox viruses, reoviruses, paramyxoviruses, rhabdoviruses, arena-viruses, and coronaviruses. Other disease states can have antigens designed for them and used in alternative embodiments of the present invention, including antigens with pathogenic protozoa, such as malaria antigen.

The fusion proteins of the present invention are preferably manufactured through expression of recombinant DNA sequences.

The DNAs used in the practice of the invention may be natural or synthetic. The recombinant DNA segments containing the nucleotide sequences coding for the embodiments of the present invention can be prepared by the following general processes:

- (a) A desired truncated gene is cut out from a plasmid in which it has been cloned, or the gene can be chemically synthesized;
- (b) An appropriate linker is added thereto as needed, followed by construction of a fused gene; and
- (c) The resulting fused protein gene is ligated down stream from a suitable promoter in an expression vector. Techniques for cleaving and ligating DNA as used in the invention are generally well known to those of ordinary skill in the art and are described in Molecular Cloning, A Laboratory Manual, (1989) Sambrook, J., et al., Cold Spring Harbor Laboratory Press.

As the promoter used in the present invention, any promoter is usable as long as the promoter is suitable for expression in the host used for the gene expression. The promoters can be prepared enzymatically from the corresponding genes, or can be chemically synthesized.

Conditions for usage of all restriction enzymes were in accordance with those of the manufacturer, including instructions as to buffers and temperatures. The enzymes were obtained from New England Biolabs, Bethesda Research Laboratories (BRL), Boehringer Mannheim and Promega.

Ligations of vector and insert DNA's were performed with T4 DNA ligase in 66mM Tris-HCl, 5mM MgCl<sub>2</sub>. ImMDTE, ImMATP, pH 7.5 at 15°C for up to 24 hours. In general, 1 to 200 ng of vector and 3-5x excess of insert DNA were preferred.

Selection of <u>E. coli</u> containing recombinant plasmids involve streaking the bacteria onto appropriate antibiotic containing LB agar plates or culturing in shaker flasks in LB liquid (Tryptone 10g/L, yeast extract 5g/L, NaCl 10g/L, pH 7.4) containing the appropriate antibiotic for selection when required. Choice of antibiotic for selection is determined by the resistance markers present on a given plasmid or vector. Preferably, vectors are selected by ampicillin.

Culturing of E. coli involves growing in Erlenmeyer flasks in LB supplemented with the appropriate anti-biotic for selection in an incubation shaker at 250-300 rpm and 37°C. Other temperature from 25°-37°C could be utilized. When cells are grown for protein production, they are induced at  $A_{560}$ =1 with IPTG to a final concentration of 0.4 mM. Other cell densities in log phase growth can alternatively be chosen for induction.

Harvesting involves recovery of E. coli cells by centrifugation. For protein production, cells are harvested 3 hours after induction though, other times of harvesting could be chosen.

In the present invention, any vector, such as a plasmid, may be used as long as it can be replicated in a procaryotic or eucaryotic cell as a host.

By using the vector containing the recombinant DNA thus constructed, the host cell is transformed via the introduction of the vector DNA.

The host cell of choice is BL21 (DE3) cells (<u>E. coli</u>), obtained from F. Wm. Studier, Brookhaven National Laboratories, Stony Brook, N.Y. Reference is also made to Wood, J. Mol. Biol., 16:118-133 (1966) U.S. Patent No. 4,952,496, and Studier, et al., J. Mol. Biol. 189:113-130 (1986). However, any strain of <u>E. coli</u> containing an IPTG inducible T7 polymerase gene would be suitable. For routine cloning, <u>E. coli</u> strain DH5α(BRL) can be used.

BL21(DE3) strain of <u>E. coli</u> was acquired under license from W. F. Studier. Reference is made to Studier, W. F. et. al., Methods in Enzymology, Vol. 185, Ch. 6, pp 60-89 (1990). This strain is unique to the extent that it contains an inducible T7 polymerase gene. The strain has no amino acid, sugar or vitamin markers, so it can grow on any rich or defined bacterial medium. It can be grown between 25°C and 37°C. It needs aeration, and it needs IPTG for induction of the T7 polymerase.

In the present invention, the fused proteins can be separated and purified by appropriate combinations of well-known separating and purifying methods. These methods include methods utilizing a solubility differential such as salt precipitation and solvent precipitation, methods mainly utilizing a difference in molecular weight such as dialysis, ultrafiltration, gel filtration and SDS-polyacrylamide gel electrophoresis, methods utilizing a difference in electric charge such as ion-exchange column chromatography, methods utilizing specific affinity such as affinity chromatography, methods utilizing a difference in hydrophobicity such as reverse-phase high pressure liquid chromatography, methods utilizing a difference in isoelectric point, such as isoelectrofusing electrophoresis, and methods using denaturation and reduction and renaturation and oxidation.

Preferred embodiments of the invention will now be described in detail in the following non-limiting examples. The most preferred embodiments of the invention are any or all of those specifically set forth in these examples. These examples are not, however, to be construed as forming the only genus that is considered as the invention, and any combination or sub-combination of the examples may themselves form a genus. These examples further illustrate details for the preparation of various embodiments of the present invention. Those skilled in the art will readily understand that known variations of the conditions and processes of the following preparative procedures can be used to prepare these embodiments.

### EXAMPLE 1

5

10

15

20

25

30

35

40

45

50

55

BS-PEMI-2

A 1.3kb Nrul/SacII fragment of plasmid pVC45-DF+T (Fig. 2) (obtained from Dr. Ira Pastan of the National Institute of Health) containing the domain I and II coding regions of Pseudomonas exotoxin (PE) (Sequence ID No. 1) was subcloned into pBluescript II SK (Stratagene, Fig. 3) restricted with HincII and SacII. The resulting construct is designated BS-PE. The influenza MI (MI) gene (Sequence ID No. 2 and 3) which codes for the matrix protein of influenza A virus was subcloned into BS-PE restricted with SacII and SacI by amplifying the MI gene from pApr701 (P. Palase, Mt. Sinai Medical Center, New York, N.Y. pApr 701 consists of the MI gene cloned into the ECORI site of pBR322, shown at Fig. 4. Reference is made to Young, J.F. et. al, Expression of Influenza Virus Genes; The Origin of Pandemic Influenza Virus; 1983) by polymerase chain reaction (PCR) (Gene Amp® PCR Reagent Kit; Perkin Elmer Cetus, Norwalk, Conn. 06859) with oligonucleotide primers which added a SacII site adjacent to MI codon number 2 (Sequence ID No. 4) and a SacI site 3' of the MI termination codon (Sequence ID No. 5). This plasmid is designated BS-PEMI-1.

The truncated ompA leader coding sequence was removed from the 5' end of the fusion gene by replacing the small Xhol/HindIII fragment of BS-PEMI-1 with the oligonucleotide sequence shown in Sequence ID No. 6. The resulting plasmid is named BS-PEMI-2 and encodes a fusion gene consisting of <u>Pseudomonas</u> exotoxin amino acids 2 through 414 joined to MI amino acids 2 to 252 (Sequence ID No. 7 and 8).

#### **EXAMPLE 2**

pVC-ompA-PEMI-2

pVC45DF+T vector was prepared by restriction digestion with HindIII and EcoRI, followed by gel purification.

The PEMI insert fragment was prepared by restriction digestion of BS-PEMI-1 with SacI, followed by T4 DNA polymerase treatment to remove the 3' overhang. EcoRI linkers were added to the blunted SacI site, followed by restriction digestion with HindIII. The HindIII-EcoRI fragment was gel purified (Molecular Cloning Manual, Gene Clean Kit, Bio 101, Inc. P.O. Box 2284, La Jolla, CA 92038) and ligated into the prepared pVC45-

DF+T vector. The resulting construct was named pVC-ompA-PEMI-2.

The ompA signal sequence was removed from the construct by restriction digestion of pVC-ompA-PEMI-2 with Xbal and HindIII. An oligonucleotide fragment containing the T7 promoter, ribosome binding site and initiation sequence was ligated into the vector whose base sequence is shown at Sequence ID No. 9. The resulting plasmid construct was named pVC-PEMI-2 and encodes a T7 polymerase-driven gene fusion consisting of PE amino acids 2 through 414 joined to influenza MI amino acids 2 through 252. The 5' and 3' ends of the coding region, as well as the PE to MI fusion site and cytotoxic T lymphocyte epitope coding sequences (Rotzschke, O. et. al., Nature 348, 252 (1990) were confirmed by DNA sequencing.

#### EXAMPLE 3

10

20

30

35

45

50

55

#### **BS-PEMa**

The influenza Ma sequence (coding for residues 57-68 of the influenza matrix protein) was obtained by amplifying a portion of the influenza M1 gene in pApr701 by polymerase chain reaction (PCR) with oligonucleotide primers which added a SacII site adjacent to influenza MI codon No. 57 (Sequence ID No. 10) and a termination codon and a SacI site 3' of the MI codon No. 68 (Sequence ID No. 11). This fragment was cut with SacII and SacI and subcloned into BS-PE digested with SacII and SacI. The resulting plasmid is named BS-PEMa-1 and was verified by sequencing through the junctions and the Ma sequence itself.

### **EXAMPLE 4**

Subcloning of PEMa from BS-PEMal into PVC45DF+T

The PEMa insert (Sequence ID No. 12) was prepared by restricting BS-PEMa-1 with SacI and removing the 3' overhang by treatment with T4 DNA polymerase, then restricting with ApaI and gel purifying.

pVC45DF + T was restricted with EcoRI and the 5' overhang filled in with Klenow enzyme treatment (Molecular Cloning Manual, ibid.). It was subsequently restricted with Apal and gel purified. The vector and fragment were ligated together, and the resulting construction was named pVC-ompA-PEMa-1. The construction was verified by sequencing across the junctions and through Ma.

The ompA leader sequence was removed from pVC-ompA-PEMa-1 by digestion with Xbal and HindIII. An oligonucleotide fragment containing the T7 promoter, ribosome binding site, initiation sequence and a build-back of the 5' end of the PE coding region (Sequence ID No. 13) was ligated to the vector. The resulting construction was named pVC-PEMa-1 and encodes a T7 polymerase driven gene fusion consisting of PE amino acids 2 to 414 joined to influenza MI amino acids 57 to 68 (Ma) Sequence ID No. 14 and 15. The 5' end of pVC-PEMa-1 was verified by sequencing through the oligonucleotide fragment.

# **EXAMPLE 5**

# 40 Construction of pVC-PEBT

A control plasmid was constructed which encodes a T7 polymerase driven gene fusion consisting of PE amino acids 2 to 414 followed by termination codons. pVC-PEMI-2 was digested with SacII and EcoRI to remove the MI sequence. The vector was gel purified and ligated to an oligonucleotide that builds back PE codon No. 414 followed by termination signals shown in Sequence ID No. 16. The resulting construction was named pVC-PEBT (Sequence ID No. 17 and 18) and was verified by sequencing across the junctions and the oligonucleotide addition.

### **EXAMPLE 6**

### **BSK-PEMI**

BSK-PEMI was made from BS-PEMI by the replacement of the 21 base pair Xhol/HindIII fragment with a 24 base pair fragment encoding a consensus eucaryotic ribosome binding site (Sequence ID No. 19). The purpose of the construct was to increase the yields of in vitro translated PEMI protein. Thus, an additional object of the invention is to increase yields of translated PEMI protein.

#### **EXAMPLE 7**

pVCPE/2 (pVC45DF+T/2)

pVCPE/2 was made by replacing the 105 base pair PpuMI/EcoRI fragment of pVC45DF+T with a 46 base pair DNA fragment encoding an inframe duplication of PE codons 604 to 613 flanked by unique cloning sites (Sequence ID No. 20). This construct is used for generating full-length molecules of PE with the deletion of residue 553 resulting in an inactivated toxin domain (sequence ID No. 21 and 22) fused to protein segments of choice between PE codons 604 and 605. One may replace the ompA signal sequence with the promoter/ribosome binding site as described for PVC-PEMI-2.

#### **EXAMPLE 8**

pVCPE/2-Ma

15

20

5

10

pVCPE/2-Ma was made by ligating into the Xmal site of pVCPE/2 a 48 base pair DNA fragment encoding amino acids 55 through 67 (Sequence ID No. 23). This construct expresses in <u>E. coli</u> full-length PE with MI amino acids 55 through 67 inserted between PE amino acid 604 and 605 (Sequence ID No. 24 and 25). One may replace the ompA signal sequence with the promoter/ribosome binding site as described for pVC-PEMI-2.

#### **EXAMPLE 9**

pVCPE/2-MI:15-106

25

30

pVCPE/2-MI:15-106 was made by subcloning a PCR-amplified DNA fragment encoding MI amino acids 15 through 106 into the XmaI site of pVCPE/2. The sequence of the oligonucleotide primers used to amplify the MI segment are those shown at Sequence ID No. 26 and 27, respectively. This construct expresses in <u>E. coli</u> full length PE with MI amino acids 15 through 106 inserted between PE amino acid 604 and 605 (Sequence ID No. 28 and 29). One may replace the ompA signal sequence with the promoter/ribosome binding site as described for pVC-PEMI-2.

#### **EXAMPLE 10**

35 pVCPEdel(403-613)

pVCPEdel(403-613) was made by restricting pVC45DF+T with SacII followed by elimination of the 3' SacII overhang with T4 DNA polymerase and the ligation of a 3-frame termination linker whose nucleic acid sequence is given at Sequence ID No. 30. This construct will express PE domains I, II and Ib only, fused to the ompA leader in E. coli.

#### **EXAMPLE 11**

pVCPEdel(403-505)

50

55

45

40

pVCPEdel(403-505) was made by restricting pVC45DF+T with SacII and XhoI followed by removal of restriction overhangs with mung bean nuclease (New England Biolabs). The vector fragment was recovered and reclosed with DNA lipase. This construct will express in <u>E. coli</u> the PE protein lacking amino acids 403 through 505.

#### EXAMPLE 12

pVCPEdel(494-505)

pVCPEdel(494-505) was made by restricting pVC45DF+T with BamHI and XhoI followed by the filling in of the 5' overhangs with Klenow fragment. The vector fragment was recovered and reclosed with DNA ligase. This construct will express in <u>E. coli</u> the PE protein lacking amino acids 494 through 505.

#### **EXAMPLE 13**

pVCPEdel(494-610)

pVCPEdel(494-610) was made by restricting PVC45DF+T with BamHI and PpuMI followed by the filling in of the 5' overhangs with Klenow fragment. The vector fragment was recovered and reclosed with DNAligase. This construct will express in <u>E. coli</u> the PE protein lacking amino acids 494 through 610. All of the pVCPEdel plasmids were useful in determining to what extent the toxin domain of PE could be truncated without resulting in the expression of an insoluble protein in <u>E. coli</u>. It thus became an additional object of the invention to provide hybrids having the minimal toxin domain of PE that would retain water solubility.

#### **EXAMPLE 14**

Addition of Sequences Between pE and MI in pVC-PEMI-2

15

20

5

10

Oligonucleotide linkers can be added at the SacII site between PE and MI in pVC-PEM-2. These linkers can be designed to add cleavage sites and/or signal sequences which can help the MI portion of the fusion protein to become available for presentation within the cell. SacII digestion cleaves the gene between the last two PE codons (for amino acids 413 and 414) and provides an appropriate site for such additions.

The following four constructions have been made by inserting linkers at the SacII site. The constructions have been verified by sequencing across the SacII junctions and through the complete linker.

### **EXAMPLE 15**

25 pVC-PE-RK-MI

This vector contains an ARG LYS(RK) cleavage site inserted into the SacII site, using an oligonucleotide linker as shown in Sequence ID No 31. The resulting amino acid sequence between amino acids 413 and 414 of PE is Gly Gly Arg Lys Ser.

30

35

55

#### **EXAMPLE 16**

pVC-PE-RKSigI-MI

This vector contains an ARG LYS(RK) cleavage site and the signal sequence that is shown in Sequence ID No. 32 from the Influenza A hemagglutinin (HA) protein inserted at the SacII site, using the oligonucleotide linker disclosed at Sequence ID No. 33. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as shown in Sequence ID No. 34.

### 40 EXAMPLE 17

PVC-PE-Sig1-MI

This vector contains the signal sequence of HA without the RK cleavage site inserted into the SacII site using the oligonucleotide linker shown at Sequence ID No. 35. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as shown at Sequence ID No. 36.

### **EXAMPLE 18**

pVC-PE-Sig2-MI

This vector contains the signal sequence shown at Sequence ID No. 37, derived from amino acids 22 to 48 from ovalbumin inserted into the SacII site, using the oligonucleotide linker of Sequence ID No. 38. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as that shown in Sequence ID No. 39.

#### Addition of Sequences Between

PE and Ma In pVC-PEMa-1

Oligonucleotide linkers can be added at the SacII site between PE and Ma in pVC-PEMa-1. These linkers can be designed to add cleavage sites and/or signal sequences which can help the Ma peptide to become available for presentation within the cell. SacII digestion cleaves the gene between the last two PE codons (for amino acids 413 and 414) and thus provides an appropriate site for such additions.

The following four examples have been made by inserting linkers at the SacII site. The constructions have been verified by sequencing across the SacII junctions and through the complete linker.

### **EXAMPLE 19**

5

10

15

20

30

35

40

pVC-PE-RKSig1-Ma

This vector contains an ARG LYS (RK) cleavage site and the signal sequence from the Influenza A hemagglutimin (HA) protein inserted into a blunted SacII site, using the oligonucleotide linker shown at Sequence ID No. 40. The resulting amino acid sequence between amino acids 413 and 414 of PE exotoxin is also as shown at Sequence ID No. 41.

### **EXAMPLE 20**

pVC-PE-Sig1-Ma

This vector contains the single sequence of HA without a cleavage site inserted into a blunted SacII site using the oligonucleotide linkers shown in Sequence ID No. 42. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as shown in Sequence ID No. 43.

### **EXAMPLE 21**

pVC-PE-Sig2-Ma

This vector contains a signal sequence derived from amino acids 22 through 48 from ovalbumin inserted into a blunted SacII site, using the oligonucleotide linker as seen in Sequence ID No. 44. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as shown in Sequence ID No. 45.

### **EXAMPLE 22**

pVC-PE-Sig1Sig2-MA

This vector contains the signal sequence derived from HA, followed by the signal sequence from ovalbumin inserted into the SacII site, using the oligonucleotide linker shown at Sequence ID No. 46. The resulting amino acid sequence between amino acids 413 and 414 of PE is also as shown at Sequence ID No. 47.

#### 45 EXAMPLE 23

BSPEMIc5aa

The plasmid BSPEMI-2 was digested with SacI and StuI and ligated to the oligonucleotide linker shown at Sequence No. 48. This linker builds back the C-terminus of the MI protein and adds the last five amino acids from the C-terminus of the PE protein, whose sequence is Arg Glu Asp Leu Lys, followed by a termination codon. This also incorporates an EcoRI site. The resulting plasmid was named BSPEMIc5aa and was sequenced across the junctions (Sequence ID No. 49 and 50) and the linker for verification of the construction.

### **EXAMPLE 24**

### pVC-PEMIc5aa

The plasmid BSPEMIc5aa was digested with HindIII and EcoRI and 1.8 kb PEMIc5aa fragment was gel purified. The plasmid pVC-PEMI-2 was digested with HindIII and EcoRI and the 3.2 kb vector fragment was ligated to the 1.8 kb PEMIc5aa fragment and the resulting plasmid was named pVC-PEMIc5aa. The 5' and 3' ends of the PEMIc5aa insert were verified by sequencing.

#### 10 EXAMPLE 25

5

15

20

30

45

55

#### pVC-PENPc5aa

A fragment containing the nucleoprotein (NP) of Influenza A virus was obtained from plasmid pApr501 (obtained from Peter Palase, Mt. Sinai Medical Center, New York, N.Y. pApr501 is said nucleoprotein gene cloned into the EcoRI site of pBR322, (Fig. 4) by polymerase chain reaction with oligonucleotide primers which added a SacII site adjacent to the ATG codon of NP to give the sequence shown at Sequence ID No. 51, and the last 5 amino acids of PE followed by a termination codon and an EcoRI site to the 3' end of NP to give the sequence shown at Sequence ID No. 52. The polymerase chain reaction fragment was digested with SacII and EcoRI and ligated to the plasmid pVC-PEMI-2 digested with SacII and EcoRI. The resulting plasmid is named pVC-PENPc5aa. The 5' and 3' ends of the PENPc5aa insert (Sequence ID No. 53 and 54) were verified by sequencing. This construction fuses the binding and translocation domains of PE to the Influenza A nucleoprotein.

#### **EXAMPLE 26**

25

pVC-ompA-PEGAG

The HIV GAG gene was obtained from plasmid HIVpBR322 (obtained from Ron Diehl Merck, Sharpe and Dohme Research Laboratories, West Point, PA., Fig. 5) by polymerase chain reaction with oligonucleotides that added a SacII site adjacent to the ATG codon of GAG to give the nucleotide sequence shown at Sequence ID No. 55, and a SacI site immediately after the termination codon at the 3' end to give the nucleotide sequence at Sequence ID No. 56. The polymerase chain reaction fragment was digested with SacII and ligated to plasmid pVC45DF+T, which had been digested with EcoRI, the 5' overhang filled in by Klenow fragment, and digested with SacII. The resulting plasmid was named pVC-ompA-PEGAG (Sequence ID No. 57 and 58) and was verified by a partial sequence at the SacII junction.

This construction fused the binding and translocation domains of PE to the GAG gene of HIV-1 virus. The fusion protein contains an ompA leader sequence. Alternatively, any vector containing the complete coding region for HIV GAG can be used with these oligomers to generate the HIV GAG gene by PCR.

#### 40 EXAMPLE 27

Expression of PEMI, PEMa and PEBT

Frozen competent BL21(DE3) cells (as described by Studier, et al. Mol. Biol., 189, 113-130, 1986) were prepared as described (DNA cloning, Vol. 1, p. 121, Ed. D N Glover, IRL Press, Wash., D.C.).

BL21(DE3) cells were transformed with pVC-PEMI-2, pVC-PEMa-1, or pVC-PEBT as described below (this can be performed with pVC-PE fusion plasmids in general) and transformants were selected on L-Amp plates. Fresh transformants were used to inoculate L-Amp liquid cultures at A560=0.1. Cultures were grown at 37°C with vigorous aeration and induced at A560=1.0 with IPTG to a final concentration of 0.4 mM. Cultures were harvested after 3 hours of induction and the cell pellets used for protein extraction and purification (Protein Structure: A Practical Approach, T.E. Creighton, ed., IRL Press at Oxford Univ. Press, Ch. 9, 191 (1989)).

#### Transformation Procedure

A bath of dry ice/ethanol was prepared and maintained at -70°C. Competent cells were removed from a -70°C freezer and thawed on ice. A sufficient number of 17 x 100 mm polypropylene tubes (Falcon 2059) were placed on ice. 100  $\mu$ l aliquots of gently mixed cells were prepared in the chilled polypropylene tubes. DNA was added by moving a pipette through the cells while dispensing; the cells were then gently shaken for 5 seconds

after addition. The cells were incubated on ice for 30 minutes, then heat-shocked in a 42°C water bath for 45 seconds without shaking. The cells were again placed on ice for 2 minutes. 0.9 ml of S.O.C. reagent (Bacto-tryptone 2%, Yeast Extract 0.5%, NaCl 10mM, KCl 2.5mM, MgCl<sub>2</sub>·MgSO<sub>4</sub> 20mM, Glucose 20mM and distilled water, up to 100 ml) was added and the mixture shaken for 1 hour at 225 rpm and 37°C, then plated on antibiotic plates, spread gently.

### **EXAMPLE 28**

10

15

20

25

30

35

40

45

50

55

Incubation of U-2 OS Cells With 51Cr and Protein/PEMa

U-2 OS cells (ATCC) were harvested from flasks, after a 1X wash with RCM 8, using 1mM EDTA. The flasks were incubated at 37°C for 10 minutes. until cells were nonadherent. Five ml. of U-2 OS medium [McCoy's 5A (GIBCO) supplemented with 15% fetal bovine serum (HyClone) and penicillin 100 U/ml and streptomycin 100  $\mu$ g/ml (GIBCO)] was added, and the cells were centrifuged for 10 minutes at 210 x g.

Cells were resuspended in U-2 OS medium at  $8.5 \times 10_s$ /ml. To each well of a 12-well plate, 0.7 ml of cell suspension was added. Negative controls include U-2 OS medium alone and PEBT. The positive control for sensitization of U-2 OS cells is KKAMI (2 µg/ml), from M. Gammon and H. Zweerink (Merck, Sharp and Dohme Research Laboratories, Rahway, NJ). PEMa was added at  $0.2 \mu$ M or greater well concentration. Simultaneously, 137.5  $\mu$ Ci of  $^{51}$ Cr (Amersham) was added to each well. Medium was added to all wells to bring the total volume to 1 ml. This was placed at 37°C, 5.5% CO<sub>2</sub> for 14 hours.

#### **EXAMPLE 29**

Assay Protocol for CTL Activity Against Sensitized U-2 OS Targets

After the 14 hour incubation, U-2 OS were removed, after a 1X RCM 8 wash using 1mM EDTA. Plates were incubated at 37°C for 10 minutes until cells were nonadherent. K medium [RPMI 1640 (GIBCO) supplemented with 10% fetal bovine serum (HyClone), 10 mM HEPES (GIBCO), 2 mM L-glutamine (GIBCO), penicillin 100 U/ml and streptomycin 100  $\mu$ g/ml (GIBCO), and 50  $\mu$ m 2-mercaptoethanol (Bio-Rad)] was added to give a total volume of 10 ml; cells were centrifuged for 10 minutes at 210 x g. The cells were incubated at room temperature for 10 minutes in 10 ml of K medium before entering the second centrifugation. The cells were then resuspended in 1 ml of K medium, counted, and resuspended to 1 x 105/ml in K medium.

Human cytotoxic T lymphocytes, generated from one donor, were harvested, centrifuged for 10 minutes at 92 x g, and resuspended in K medium at 2.5 x 10<sup>6</sup>/ml.

100  $\mu$ l of human CTLs were added to each well of a 96-well U-bottom microtiter plate (CoStar). 100  $\mu$ l of the U-2 OS  $^{51}$ Cr-labeled targets were also added to these wells for a final effector/target ratio of 25:1. Spontaneous  $^{51}$ Cr release was determined by incubating U-2 OS cells with 100  $\mu$ l of K medium alone. The maximal release was determined by adding 100  $\mu$ l of 6 M HCl to 100  $\mu$ l of targets. The plates were quickly centrifuged to bring down the cells, and incubated for 2 hours at 37°C.

After this 2 hour incubation, the plates were centrifuged for 5 minutes, 330 x g,  $5^{\circ}$ C; 30  $\mu$ l of supernatant was harvested from each well onto a plastic-backed filtermat (Pharmacia/LKB). The mat was dried in the microwave for 3 minutes. on medium-high power. The mat was placed into a sample bag with 10 ml of BetaPlate Scint, heat sealed and placed into the BetaPlate 1205 counter (Pharmacia/LKB). Results were expressed as % specific lysis, defined as:

where

Experimental = counts per minute from the 30  $\mu$ l of supernatant harvested from the wells containing targets plus human cytotoxic T lymphocytes, as determined by a BetaPlate 1205 counter;

Spontaneous = counts per minute from the 30  $\mu$ l of supernatant harvested from the wells containing targets plus medium alone, as determined by the BetaPlate 1205 counter; and

Maximal = counts per minute from the 30 μl of supernatant harvested from the wells containing target plus 6M HCI (Fisher Scientific), as determined by the BetaPlate 1205 counter.

Results are presented graphically in Fig. 5, with U-2 OS medium alone and PEBT as negative controls, and KKAMI as a positive control. Greater that 10% specific lysis is considered a positive response (Cerottini, et.al., J. Exp. Med. 140:703, 1974).

### **EXAMPLE 30**

5

10

15

20

25

35

40

45

Generation of MI-specific Human Cytotoxic T Lymphocytes

Original stock of human cytotoxic T lymphocytes was derived by harvesting blood from one donor into a syringe (Becton Dickinson) containing 25 U of heparin for each ml of whole blood (Elkins-Sinn, Inc.). The heparinized blood was pipetted directly into a Leucoprep tube (Becton Dickinson) and centrifuged for 20 minutes at 1700 X g. The buffy coat which was seen just above the interface was removed, centrifuged for 10 minutes at 92 X g, and washed twice in RPMI 1640 (GIBCO). The peripheral blood mononuclear cells (PBLs) recovered from the Leucoprep procedure were resuspended in 10 ml of CTL medium [RPMI 1640 (GIBCO) supplemented with 10% donor or pooled human plasma, 4 mM L-glutamine, 10 mM HEPES, penicillin 100 U/ml and streptomycin 100 µg/ml (GIBCO)] at 1 X 10<sup>8</sup>/ml.

MI peptide (received from M. Gammon and H. Zweerink, MSDRL, Rahway; 2 mg/ml stock) in DMSO was diluted 1:10 in RPMI 1640 (GIBCO). MI peptide was added to the 10 ml of lymphocytes at a final concentration of 5 μg/ml. The cells were then plated at 1.5 X 10<sup>6</sup>/well in 24-well plates (Nunc).

Two U/ml of Interleukin-2 ala-125 (Amgen) was added on Day 3. The cell density was adjusted to 1 X 10<sup>6</sup>/ml as needed, and the medium was supplemented with 2 U/ml additional Interleukin-2 to compensate for the increase in volume. Cells were restimulated with peptide-pulsed peripheral blood lymphocytes every 7 days as described below. Interleukin-2 ala-125 (Amgen) was replenished every 3 days.

Cytotoxic T lymphocytes and unstimulated PBLs were frozen (CryoMed) in a mixture of 70% RPMI 1640 (GIBCO), 20% fetal bovine serum (HyClone), and 10% dimethyl sulfoxide (Sigma) and thawed as needed.

#### EXAMPLE 31

Recovery and Restimulation of Frozen CTL's

Cytotoxic T lymphocytes (CTL's) were thawed in a 37° water bath and then resuspended in 35 ml of CTL medium [RPMI 1640 (GIBCO) supplemented with 10% donor or pooled human plasma, 4 mM L-glutamine, 10 mM HEPES, penicillin 100 U/ml and streptomycin 100  $\mu$ g/ml (GIBCO]. The cytotoxic T lymphocytes were then placed at 37°, 5% CO<sub>2</sub> for 1 hour. The cell suspension was centrifuged for 10 minutes at 92 X g. The cells were resuspended at 5 X 10<sup>5</sup>/ml in CTL medium.

The source of stimulator cells for the freshly thawed cytotoxic T lymphocytes was freshly harvested PBL, which had been collected using the Leucoprep method described above. For peptide pulsing, an appropriate number (2 x 10<sup>6</sup> - 10<sup>7</sup>) of PBL were centrifuged, the supernatant was aspirated, and KKAMI at 200 μg/ml in RPMI 1640 (GIBCO) plus 10% DMSO (Sigma) was added at the rate of 100 μl of KKAMI for every 10<sup>7</sup> cells. The cells were incubated for 1 hour at 37°, 5% CO<sub>2</sub>. The peptide-pulsed peripheral blood lymphocytes were irradiated with 2,000 Rads using a <sup>60</sup>Co source. The cells were washed once in RPMI 1640, centrifuged for 10 minutes at 92 X g, and resuspended in CTL medium at 1 X 10<sup>6</sup>/ml.

": .

1.

Equal volumes of cytotoxic T lymphocytes and irradiated, peptide-pulsed peripheral blood lympocytes were mixed together for a final ratio of 1 CTL:2 peptide-pulsed PBL. Interleukin-2 ala-125 (Amgen) was added at a final concentration of 2 U/ml. The cells were thoroughly mixed together with the Interleukin-2 ala-125 and 1.2 ml was plated into each well of a 48-well plate (CoStar).

The cells were counted and Interleukin-2 ala-125 was replenished every 3 days. This was achieved by pooling all the wells into a centrifuge tube, counting the cells in a hemocytometer counting chamber, adjusting the cells to 1 X 10<sup>6</sup>/ml with CTL medium, and adding 2 U/ml of Interleukin-2 ala-125. Then 1.5 X 10<sup>6</sup> cytotoxic Tlymphocytes in 1.5 ml of CTL medium with Interleukin-2 ala-125 were plated into each well of a 24-well plate (CoStar), the restimulation process was repeated every seven days, at which time frozen PBL's were then used as the source of stimulators.

#### 50 Example 32

Binding of PEMa to the PE receptor

PEMa was used in a binding/competition assay to compete with PE for the PE receptor on U-2 OS cells. In doing so, PEMa was shown in Figure 6 to protect the cells from the toxic effects of PE. Therefore, replacement of the toxin domain of PE with the Influenza matrix peptide (amino acids 57-68) did not prohibit the binding of this chimeric protein to the PE receptor. This suggests that the ability of PEMa to sensitize target cells for lysis by CTLs specific for the matrix peptide is mediated through PE receptor-mediated uptake and processing.

U-2 cells were grown to a density of 20,000 cells/100μl in 960 well plates. Cells were preincubated with PEMA (0,0.1, 1, 10 and 50 μg in 100 μl of complete McCoy's 5A medium) for 30 minutes at 37°C, followed by incubation with or without PE(10 ng) for 2 minutes. This represents a 0-, 10-, 100-, 1000-, and 5000-fold excess of PEMA over PE, respectively. Cells were washed with McCoy's medium (3 x 200 μl); then incubated with [35S]methionine (2 μCi/100 μl) for an additional 5 hours at 37°C and washed (3 x 200 μl). Cells were lysed in 10mM EDTA (100 μl) and aliquots (5 μl) were spotted onto Whatman 3MM filters. Incorporation of radioactivity was assayed by TCA precipitation of the cellular proteins onto the filter papers by immersion into ice-cold TCA (10% w/v) for at least 1 hour. Filters were washed once with 5% TCA and 3 times with ethanol and dried. Radioactivity was determined by liquid scintillation counting. Incorporation of [35S]methionine into the TCA-precipitable pool of cellular proteins in the absence (open circles) or presence (closed circles) of PE is shown as a function of lop excess PEMa. Error bars represent +/-SEM for n=9. Using a one-tailed t-test, incorporation of [35S]methionine was determined to be significantly lower in the presence of PE than in the absence of PE at 0-, 10-, and 100-fold excesses of PEMa (99.5%, 99.5% and 95% confidence limits, respectively). However, at 1000- and 5000-fold excesses of PEMa, incorporation was not significantly different in the presence or absence of PE.

Following preparation of the protein hybrids of the present invention, a suspension of the protein-hybrids suitable for injection into the host animal must be prepared. Typical suspension vehicles include sterile saline and sterile water for injection. Various agents may be added as preservatives including benzethonium chloride (0.0025%), phenol (0.5%), thiomersal (1:10,000). Strength of the vaccine will be measured as mass of fusion protein which generates a protective response, defined by in vitro/in vivo results, per given host species, a method known to those of ordinary skill in the art.

The suspensions for injection must, of course, be prepared under sterile conditions, in which there is a total absence of living organisms and absolute freedom from biological contamination present in the suspension for injection.

Although water is always the solvent of choice for an injectable preparation, co-solvents that may be additionally present include ethyl alcohol, glycerin, propylene glycol, polyethylene glycol and dimethylacetamide. Buffers may be added, including acidic acid, citric acid or phosphoric acid systems. Antioxidants can include ascorbic acid, BHA, BHT, sodium bisulfite, and sodium metabisulfite. Tonicity can be adjusted with agents such as dextrose, sodium chloride and sodium sulfate.

Aseptic manufacture of vaccines, including their packaging, is conducted according to methods well known to those of ordinary skill in the art, and as described in standard texts on the subject, including Lachman, L., et al., <a href="https://example.com/herrican-pharmacy">The Theory And Practice of Industrial Pharmacy</a>, Dittert, L., ed, <a href="https://example.com/sprows/sprows-pharmacy">Sprows/s American Pharmacy</a>; and <a href="https://example.com/sprows-pharmacy">Re-mington's Pharmaceutical Sciences</a>.

While the invention has been described and illustrated in reference to certain preferred embodiments thereof, those skilled in the art will appreciate that various changes, modifications and substitutions can be made therein without departing from the spirit and scope of the invention. It is intended, therefore, that the invention be limited only by the scope of the claims which follow, and that such claims be interpreted as broadly as is reasonable.

SEQUENCE LISTING

45

10

15

20

25

30

35

40

50

# (2) INFORMATION FOR SEQ ID NO:1:

(3)	SEQUENCE	CHARACTERISTICS:
-----	----------	------------------

(A) LENGTH: 1294 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

15	TCGCGATTGC AGTGGCACTG GCTGGTTTCG CTACCGTAGC GCAGGCCGCG AATTTGGCCG	60
	AAGAAGCTTT CGACCTCTGG AACGAATGCG CCAAAGCCTG CGTGCTCGAC CTCAAGGACG	120
20	GCGTGCGTTC CAGCCGCATG AGCGTCGACC CGGCCATCGC CGACACCAAC GGCCAGGGCG	180
	TGCTGCACTA CTCCATGGTC CTGGAGGGCG GCAACGACGC GCTCAAGCTG GCCATCGACA	240
	ACGCCCTCAG CATCACCAGC GACGGCCTGA CCATCCGCCT CGAAGGCGGC GTCGAGCCGA	300
25	ACAAGCCGGT GCGCTACAGC TACACGCGCC AGGCGCGCGG CAGTTGGTCG CTGAACTGGC	360
	TGGTACCGAT CGGCCACGAG AAGCCCTCGA ACATCAAGGT GTTCATCCAC GAACTGAACG	420
30	CCGGCAACCA GCTCAGCCAC ATGTCGCCGA TCTACACCAT CGAGATGGGC GACGAGTTGC	480
	TGGCGAAGCT GGCGCGCGAT GCCACCTTCT TCGTCAGGGC GCACGAGAGC AACGAGATGC	540
	AGCCGACGCT CGCCATCAGC CATGCCGGGG TCAGCGTGGT CATGGCCCAG ACCCAGCCGC	600
35	GCCGGGAAAA GCGCTGGAGC GAATGGGCCA GCGGCAAGGT GTTGTGCCTG CTCGACCCGC	660
	TGGACGGGGT CTACAACTAC CTCGCCCAGC AACGCTGCAA CCTCGACGAT ACCTGGGAAG	720
40	GCAAGATCTA CCGGGTGCTC GCCGGCAACC CGGCGAAGCA TGACCTGGAC ATCAAACCCA	780
	CGGTCATCAG TCATCGCCTG CACTTTCCCG AGGGCGGCAG CCTGGCCGCG CTGACCGCGC	840
	ACCAGGCTTG CCACCTGCCG CTGGAGACTT TCACCCGTCA TCGCCAGCCG CGCGGCTGGG	900
45	AACAACTGGA GCAGTGCGGC TATCCGGTGC AGCGGCTGGT CGCCCTCTAC ETGGCGGCGC	960
	COCTOTOCTO CAACCAGGIC GACCAGGIGA TOOGCAACGO COTGGOCAGGO COCGGOCAGGG	1020

50

5

10

	GCGGCGACCT GGGCGAAGCG ATCCGCGAGC AGCCGGAGCA GGCCCGTCTG GCCCTGACCC	1080										
	TGGCCGCCGC CGAGAGCGAG CGCTTCGTCC GGCAGGGCAC CGGCAACGAC GAGGCCGGCG	1140										
5	CGGCCAACGC CGACGTGGTG AGCCTGACCT GCCCGGTCGC CGCCGGTGAA TGCGCGGGCC	1200										
	CGGCGGACAG CGGCGACGCC CTGCTGGAGC GCAACTATCC CACTGGCGCG GAGTTCCTCG	1260										
10	GCGACGGCGG CGACGTCAGC TTCAGCACCC GCGG	1294										
	(2) INFORMATION FOR SEQ ID NO:2:											
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 759 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear											
20	(ii) MOLECULE TYPE: DNA (genomic)											
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:											
25	ATGAGTETTE TAACEGAGGT CGAAACGTAC GTTETETTA TEATECEGTE AGGECECETE	. 60										
	AAAGCCGAGA TCGCACAGAG ACTTGAAGAT GTCTTTGCAG GGAAGAACAC CGATCTTGAG	120										
30	GTTCTCATGG AATGGCTAAA GACAAGACCA ATCCTGTCAC CTCTGACTAA GGGGATTTTA	180										
	GGATTTGTGT TCACGCTCAC CGTGCCCAGT GAGCGAGGAC TGCAGCGTAG ACGCTTTGTC	24										
35	CAAAATGCCC TTAATGGGAA CGGGGATCCA AATAACATGG ACAAAGCAGT TAAACTGTAT	300										
	AGGAAGCTCA AGAGGGAGAT AACATTCCAT GGGGCCAAAG AAATCTCACT CAGTTATTCT	360										
	GCTGGTGCAC TTGCCAGTTG TATGGGCCTC ATATACAACA GGATGGGGGC TGTGACCACT	420										
40	GAAGTGGCAT TTGGCCTGGT ATGTGCAACC TGTGAACAGA TTGCTGACTC CCAGCATCGG	480										
	TCTCATAGGC AAATGGTGAC AACAACCAAC CCACTAATCA GACATGAGAA CAGAATGGTT	540										
45	TTAGCCAGCA CTACAGCTAA GGCTATGGAG CAAATGGCTG GATCGAGTGA GCAAGCAGCA	601										
	GAGGCCATGG AGGTTGCTAG TCAGGCTAGG CAAATGGTGC AAGCGATGAG AACCATTGGG	660										
50												

	ACTCATCCTA GCTCCAGTGC TGGTCTGAAA AATGATCTTC TTGAAAATTT GCAGGCCTAT	720
5	CAGAAACGAA TGGGGGTGCA GATGCAACGG TTCAAGTGA	759
	(2) INFORMATION FOR SEQ ID NO:3:	
10	<ul> <li>(i) SEQUENCE CHARACTERISTICS:         <ul> <li>(A) LENGTH: 253 amino acids</li> <li>(B) TYPE: amino acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul> </li> </ul>	
15	(ii) MOLECULE TYPE: protein	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
20	Met Ser Leu Leu Thr Glu Val Glu Thr Tyr Val Leu Ser Ile Ile Pro	
25	Ser Gly Pro Leu Lys Ala Glu Ile Ala Gln Arg Leu Glu Asp Val Phe 20 25 30	
25	Ala Gly Lys Asn Thr Asp Leu Glu Val Leu Met Glu Trp Leu Łys Thr 35 40 45	
30	Arg Pro Ile Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe 50 55 60	
	Thr Leu Thr Val Pro Ser Glu Arg Gly Leu Gln Arg Arg Arg Phe Val 65 70 75 80	
35	Gln Asn Ala Leu Asn Gly Asn Gly Asp Pro Asn Asn Met Asp Lys Ala 85 90 95	
	Val Lys Leu Tyr Arg Lys Leu Lys Arg Glu Ile Thr Phe His Gly Ala 100 105 110	
40	Lys Glu Ile Ser Leu Ser Tyr Ser Ala Gly Ala Leu Ala Ser Cys Met 115 120 125	
45	Gly Leu Ile Tyr Asn Arg Met Gly Ala Val Thr Thr Glu Val Ala Phe 130 135 140	
	Gly Leu Val Cys Ala Thr Cys Glu Gln Ile Ala Asp Ser Gln His Arg 145 150 155 160	
50		

	Ser His Arg Gln Met Val Thr Thr Asn Pro Leu Ile Arg His Glu 165 170 175	
5	Asn Arg Met Val Leu Ala Ser Thr Thr Ala Lys Ala Met Glu Gln Met 180 185 190	
10	Ala Gly Ser Ser Glu Gln Ala Ala Glu Ala Met Glu Val Ala Ser Gln 195 200 205	
	Ala Arg Gln Met Val Gln Ala Met Arg Thr Ile Gly Thr His Pro Ser 210 215 220	
15	Ser Ser Ala Gly Leu Lys Asn Asp Leu Leu Glu Asn Leu Gln Ala Tyr 225 230 235 240	
	Gln Lys Arg Met Gly Val Gln Met Gln Arg Phe Lys Xaa 245 250	
20	(2) INFORMATION FOR SEQ ID NO:4:  (i) SEQUENCE CHARACTERISTICS:	
25	(1) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 30 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:	
	ATACCCGCGG CAGTCTTCTA ACCGAGGTCG	30
35	(2) INFORMATION FOR SEQ ID NO:5:	
40	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 36 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
15	(ii) MOLECULE TYPE: DNA (genomic)	
45	() CECHENCE DESCRIPTION, SEC ID NO.E.	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  CCCCACGTCT ACGTTGCCAA GTTCACTCTC GAGATA	36
50		-

	(2) INFURMATION FOR SEQ TO NO.0.	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 27 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	27
	CTCGAGAATT CATGGCCGAG GAAGCTT	27
	(2) INFORMATION FOR SEQ ID NO:7:	
20	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1998 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25	(II) TOPOLOGY. Timear	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:	
		60
	ATGGCCGAAG AAGCTTTCGA CCTCTGGAAC GAATGCGCCA AAGCCTGCGT GCTCGACCTC	00
	AAGGACGGCG TGCGTTCCAG CCGCATGAGC GTCGACCCGG CCATCGCCGA CACCAACGGC	120
35		
33	CAGGGCGTGC TGCACTACTC CATGGTCCTG GAGGGCGGCA ACGACGCGCT CAAGCTGGCC	180
		240
	ATCGACAACG CCCTCAGCAT CACCAGCGAC GGCCTGACCA TCCGCCTCGA AGGCGGCGTC	240
	GAGCCGAACA AGCCGGTGCG CTACAGCTAC ACGCGCCAGG CGCGCGGCAG TTGGTCGCTG	300
40		
	AACTGGCTGG TACCGATCGG CCACGAGAAG CCCTCGAACA TCAAGGTGTT CATCCACGAA	360
		426
	CTGAACGCCG GCAACCAGCT CAGCCACATG TCGCCGATCT ACACCATCGA GATGGGCGAC	420
45	GAGTIGCTGG CGAAGCTGGC GCGCGATGCC ACCTTCTTCG TCAGGGCGCA CGAGAGCAAC	480
	GAGTIGETEG CHARGETEGE GEGEGATEGE ACCTIONED PERSONNEL	
	GAGATGCAGC CGACGCTCGC CATCAGCCAT GCCGGGGTCA GCGTGGTCAT GGCCCAGACC	540
50		
50		

	CAGCCGCGCC	GGGAAAAGCG	CTGGAGCGAA	TGGGCCAGCG	GCAAGGTGTT	GTGCCTGCTC	600
	GACCCGCTGG	ACGGGGTCTA	CAACTACCTC	GCCCAGCAAC	GCTGCAACCT	CGACGATACC	660
5	TGGGAAGGCA	. AGATCTACCG	GGTGCTCGCC	GGCAACCCGG	CGAAGCATGA	CCTGGACATC	720
	AAACCCACGG	TCATCAGTCA	TCGCCTGCAC	TTTCCCGAGG	GCGGCAGCCT	GGCCGCGCTG	780
10	ACCGCGCACC	AGGCTTGCCA	сствссвств	GAGACTTTCA	CCCGTCATCG	CCAGCCGCGC	840
	GGCTGGGAAC	AACTGGAGCA	GTGCGGCTAT	CCGGTGCAGC	GGCTGGTCGC	CCTCTACCTG	900
	GCGGCGCGGC	TGTCGTGGAA	CCAGGTCGAC	CAGGTGATCC	GCAACGCCCT	GGCCAGCCCC	960
15	GGCAGCGGCG	GCGACCTGGG	CGAAGCGATC	CGCGAGCAGC	CGGAGCAGGC	CCGTCTGGCC	1020
	CTGACCCTGG	CCGCCGCCGA	GAGCGAGCGC	TTCGTCCGGC	AGGGCACCGG	CAACGACGAG	1080
20	ecceecece	CCAACGCCGA	CGTGGTGAGC	CTGACCTGCC	CGGTCGCCGC	CGGTGAATGC	1140
	eceeeccee	CGGACAGCGG	CGACGCCCTG	CTGGAGCGCA	ACTATCCCAC	TGGCGCGGAG	1200
	TTCCTCGGCG	ACGGCGGCGA	CGTCAGCTTC	AGCACCCGCG	GCAGTCTTCT	AACCGAGGTC	1260
25	GAAACGTACG	TTCTCTCTAT	CATCCCGTCA	GGCCCCCTCA	AAGCCGAGAT	CGCACAGAGA	1320
	CTTGAAGATG	TCTTTGCAGG	GAAGAACACC	GATCTTGAGG	TTCTCATGGA	ATGGCTAAAG	. 1380
30	ACAAGACCAA	TCCTGTCACC	TCTGACTAAG	GGGATTTTAG	GATTTGTGTT	CACGCTCACC	1440
	GTGCCCAGTG	AGCGAGGACT	GCAGCGTAGA	CGCTTTGTCC	AAAATGCCCT	TAATGGGAAC	1500
	GGGGATCCAA	ATAACATGGA	CAAAGCAGTT	AAACTGTATA	GGAAGCTCAA	GAGGGAGATA	1560
35	ACATTCCATG	GGGCCAAAGA	AATCTCACTC	AGTTATTCTG	CTGGTGCACT	TGCCAGTTGT	1620
	ATGGGCCTCA	TATACAACAG	GATGGGGGCT	GTGACCACTG	AAGTGGCATT	TGGCCTGGTA	1680
40	TGTGCAACCT	GTGAACAGAT	TGCTGACTCC	CAGCATCGGT	CTCATAGGCA	AATGGTGACA	1740
	ACAACCAACC	CACTAATCAG	ACATGAGAAC	AGAATGGTTT	TAGCCAGCAC	TACAGCTAAG	1800
	GCTATGGAGC	AAATGGCTGG	ATCGAGTGAG	CAAGCAGCAG	AGGCCATGGA	GGTTGCTAGT	1860
45	CAGGCTAGGC	AAATGGTGCA	AGCGATGAGA	ACCATTGGGA	CTCATCCTAG	CTCCAGTGCT	1920
	GGTCTGAAAA	ATGATCTTCT	TGAAAATTTG	CAGGCCTATC	AGAAACGAAT	GGGGGTGCAG	1980
	ATCCAACCCT	TCAACTCA					1005

	(2) I	NFOR	MATI	ON F	OR S	EQ I	D NO	:8:									
5		(i)	(A) (B) (C)	ENCE LEN TYP STR TOP	GTH: E: a ANDE	666 mino DNES	ami aci S: s	no a d ingl	cids	-							
10	. (	ii)	MOLE	CULE	TYP	E: p	rote	in									
	(	×i)	SEQU	ENCE	DES	CRIP	TION	: SE	Q IO	NO:	8:						
15		Met 1	Ala	G1u <sub>.</sub>	Glu	A1a 5	Phe	Asp	Leu	Trp	Asn 10	Gl u	Cys	Ala ·	Lys	Ala 15	Cys
20		Val	Leu	Asp	Leu 20	Lys	Asp	G1 y	Val	Arg 25	Ser	Ser	Arg	Met	Ser 30	Val	Asp
		Pro	Ala	Ile 35	Ala	Asp	Thr	Asn	G1 y 40	Gln	G1 y	Val	Leu	His 45	Tyr	Ser	Het
25		Va1	Leu 50	Glυ	G1 y	G1 y	Asn	Asp 55	Ala	Leu	Lys	Leu	A1 a 60	Ile	Asp	Asn	Ala
		Leu 65	Ser	Ile	Thr	Ser	Asp 70	G1 y	Leu	Thr	Ile	Arg 75	Leu	G1 u	G1 y	Gly	Va 1 80
30		Glu	Pro	Asn	Lys	Pro 85	Val	Arg	Tyr	Ser	Tyr 90	Thr	Arg	Gln	Ala	Arg 95	G1 y
35		Ser	Trp	Ser	Leu 100	Asn	Trp	Leu	Val	Pro 105	Ile	Gly	His	GΊυ	Lys 110	Pro	Ser
		Asn	Ile	Lys 115	Val	Phe	Ile	His	G1 u 120	Leu	Asn	Ala	G1 y	Asn 125		Leu	Ser
40		His	Met 130	Ser	Pro	Ile	Tyr	Thr 135		Glu	Het	Gly	Asp 140		Leu	Leu	Ala
		Lys 145		Ala	Arg	Asp	A1a 150		Phe	Phe	Val	Arg 155		His	G1 u	Ser	Asn 160
45		G1 u	Met	Gln	Pro	Thr 165		Ala	Ile	Ser	His 170		G1 y	Val	Ser	Va1 175	Val
		Met	. Ala	G1n	Thr 180		Pro	Arg	Arg	185		Arg	Trp	Ser	G1u 190		Ala
50																	

	Ser	G1 y	Lys 195	Val	Leu	Cys	Leu	Leu 200	Asp	Pro	Leu	Asp	G1 y 205	Val	Tyr	Asn
5	Tyr	Leu 210	Ala	Gln	Gln	Arg	Cys 215	Asn	Leu	Asp	Asp	Thr 220	Trp	Glu	Gly	Lys
	I1e 225	Tyr	Arg	Val	Leu	A1a 230	G1 y	Asn	Pro	Ala.	Lys 235	His	Asp	Leu	Asp	Ile 240
10	Lys	Pro	Thr	Val	Ile 245	Ser	His	Arg	Leu	His 250	Phe	Pro	Glu	G1 y	G1 y 255	Ser
15	Leu	Αla	Ala	Lev 260	Thr	Ala	His	G1 n	A1a 265	Cys	His	Leu	Pro	Leu 270	GΊυ	Thr
	Phe	Thr	Arg 275	His	·Arg	Gln	Pro	Arg 280	G1 y	Trp	Glu	-G1 n	Leu 285	G1 u	G1 n	Cys
20	Gly	Tyr 290	Pro	Va1	Gln	Arg	Leu 2 <b>9</b> 5	Val	Ala	Lev	Tyr	Leu 300	Ala	Ala	Arg	Leu
	Ser 305	Trp	Asn	Gln	Val	Asp 310	Gln	·Va 1	Ile	Arg	Asn 315	Ala	Leu	Ala	Ser	Pro 320
25	Gly	Ser	Gly	Gly	Asp 325	Leu	G1 y	Ģīu	Ala	I1e 330	Arg	Glu	Gln	Pro	G1u 335	Gln
30	Ala	Arg	Leu	Ala 340	Leu	Thr	Leu	Ala	A1a 345	Ala	Glu	Ser	Glu	Arg 350	Phe	Val
	Arg	, G1 n	G1 y 355	Thr	G1 y	Asn	Asp	G1 ս 360		G1 y	Ala	Ala	Asn 365		Asp	Val
35	Val	Ser 370		Thr	Cys	Pro	Va1 375		Ala	G1 y	Glu	Cys 380		G1 y	Pro	Ala
	Asp 385		Gly	Asp	Ala	<b>Leu</b> 390		Glu	Arg	Asn	Tyr 395		Thr	Gly	Ala	G1 u 400
40	Phe	Leu	G1 y						Ser			Thr	Arg	G1 y	Ser 415	Leu
45	Leu	Thr	Glu	Va1 420		Thr	Tyr	Val	Leu 425		Ile	Ile	Pro	Ser 430		Pro
45	Leu	Lys	A1 a 435		Ile	Ala	Glm	440		G1u	Asp	Val	Phe 445		G1y	Lys

	Asņ	Thr 450	Asp	Leu	GΊυ	Val	Leu 455	Het	Glu	Trp	Leu	Lys 460	Thr	Arg	Pro	Ile
5	Leu 465	Ser	Pro	Leu	Thr	Lys 470	G1 y	Ile	Leu	G1 y	Phe 475	Val	Phe	Thr	Lev	Thr 480
	Val	Pro	Ser	G1u	Arg 485	Gly	Leu	G1n	Arg	Arg 490	Arg	Phe	Val	Gln	Asn 495	Ala
10	Lev	Asn	Gly	Asn 500	Gly	Asp	Pro	Asn	Asn 505	Met	Asp	Lys	Ala	Val 510	Lys	Leu
15	Tyr	Arg	Lys 515	Lev	Lys	Arg	G1 u	Ile 520	Thr	Phe	His	Gly	A1a 525	Lys	Glu	Ile
	Ser	Leu 530	Ser	Tyr	Ser	Ala	G1 y 535	Ala	Leu	Ala	"Ser	Cys 540	Met	G1 y	Leu	Ile
20	Tyr . 545	Asn	Arg	Met	61 y	A1a 550	Va1	Thr	Thr	G1 u	Va1 555		Phe	G1 y	Leu	Va1 560
	Cys	Ala	Thr	Cys	G1 u 565	Gln	Пe	Ala	Asp	Ser 570	Gln	His	Arg	Ser	Hi s 575	Arg
25	Gln	Met	Val	Thr 580	Thr	Thr	Asn	Pro	Leu 585	Ile	Arg	His	Glu	Asn 590	Arg	Met
30	Val	Leu	A1a 595	Ser	Thr	Thr	Ala	Lys 600	Ala	Met	Glu	G1n	Met 605	Ala	Gly	Ser
	Ser	G1 u 610	Gln	Ala	Ala	GΊυ	A1a 615	Met	G1 u	Val	Ala	Ser 620	G1 n	Ala	Arg	Gln
35	Met 625	Val	Gln	Ala	Met	Arg 630	Thr	Ile	G1 y	Thr	His 635		Ser	Ser	Ser	A1a 640
	Gly	Leu	Lys	Asn	Asp 645	Leu	Leu	G1 u	Asn	Leu 650		Ala	Tyr	Gln	Lys 655	Arg
40	Met	Gly	۷a۱	G1n 660	Met	G1 n	Arg	Phe	Lys 665	Xaa						
(2	) INFO	RMAT	ION	FOR	SEQ	ID N	0:9:									
45	(i)	(A (B (C	UENC ) LE ) TY ) ST ) TO	NGTH PE: RAND	: 52 nucl EDNE	bas eic SS:	e pa acid sing	irs								
50										ě						

(ii) MOLECULE TYPE: DNA (genomic)

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
	CTAGAAATAA TTTTGTTTAA CTTTAAGAAG GAGATATACA TATGGCCGAA GA	52
10	(2) INFORMATION FOR SEQ ID NO:10:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 32 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
15	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
20		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
	ATACCCGCGG CAAGGGGATT ITAGGATTIG IG	32
25	(2) INFORMATION FOR SEQ ID NO:11:	·
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 36 base pairs	
	(B) TYPE: nucleic acid	
30	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
35		
35		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
	ATAGAGCTCT CACACGGTGA GCGTGAACAC AAATCC	36
40	(2) INFORMATION FOR SEQ ID NO:12:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 52 base pairs	
45	(B) TYPE: nucleic acid	
45	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	

50

55

(ii) MOLECULE TYPE: DNA (genomic)

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
	CCGCGGCAAG GGGATTTTAG GATTTGTGTT CACGCTCACC GTGTGAGAGC TC	52
10	(2) INFORMATION FOR SEQ ID NO:13:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 52 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	
15	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	•	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
	CTAGAAATAA TTTTGTTTAA CTTTAAGAAG GAGATATACA TATGGCCGAA GA	52
25	(2) INFORMATION FOR SEQ ID NO:14:	
30	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 1281 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
	(ii) MOLECULE TYPE: DNA (genomic)	
35		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
	ATGGCCGAGG AAGCTTTCGA CCTCTGGAAC GAATGCGCCA AAGCCTGCGT GCTCGACCTC	60
40	AAGGACGGCG TGCGTTCCAG CCGCATGAGC GTCGACCCGG CCATCGCCGA CACCAACGGC	120
	CAGGGCGTGC TGCACTACTC CATGGTCCTG GAGGGCGGCA ACGACGCGCT CAAGCTGGCC	180
45	ATCGACAACG CCCTCAGCAT CACCAGCGAC GGCCTGACCA TCCGCCTCGA AGGCGGCGTC	240
	GAGCCGAACA AGCCGGTGCG CTACAGCTAC ACGCGCCAGG CGCGCGGCAG TTGGTCGCTG	300
50		

25

AACTGGCTGG	TACCGATCGG	CCACGAGAAG	CCCTCGAACA	TCAAGGTGTT	CATCCACGAA	360
CTGAACGCCG	GCAACCAGCT	CAGCCACATG	TCGCCGATCT	ACACCATCGA	GATGGGCGAC	420
GAGTTGCTGG	CGAAGCTGGC	GCGCGATGCC	ACCTTCTTCG	TCAGGGCGCA	CGAGAGCAAC	480
GAGATGCAGC	CGACGCTCGC	CATCAGCCAT	GCCGGGGTCA	GCGTGGTCAT	GGCCCAGACC	540
CAGCCGCGCC	GGGAAAAGCG	CTGGAGCGAA	TGGGCCAGCG	GCAAGGTGTT	GTGCCTGCTC	600
GACCCGCTGG	ACGGGGTCTA	CAACTACCTC	GCCCAGCAAC	GCTGCAACCT	CGACGATACC	660
TGGGAAGGCA	AGATCTACCG	GGTGCTCGCC	GGCAACCCGG	CGAAGCATGA	CCTGGACATC	720
AAACCCACGG	TCATCAGTCA	TCGCCTGCAC	TTTCCCGAGG	GCGGCAGCCT	GGCCGCGCTG	780
ACCGCGCACC	AGGCTTGCCA	CCTGCCGCTG	GAGACTTTCA	CCCGTCATCG	CCAGCCGCGC	840
GGCTGGGAAC	AACTGGAGCA	GTGCGGCTAT	CCGGTGCAGC	GGCTGGTCGC	CCTCTACCTG	900
ececcecec	TGTCGTGGAA	CCAGGTCGAC	CAGGTGATCC	GCAACGCCCT	GGCCAGCCCC	960
GGCAGCGGCG	GCGACCTGGG	CGAAGCGATC	CGCGAGCAGC	CGGAGCAGGC	CCGTCTGGCC	1020
CTGACCCTGG	CCGCCGCCGA	GAGCGAGCGC	TTCGTCCGGC	AGGGCACCGG	CAACGACGAG	1080
GCCGGCGCGG	CCAACGCCGA	CGTGGTGAGC	CTGACCTGCC	CGGTCGCCGC	CGGTGAATGC	1140
GCGGGCCCGG	CGGACAGCGG	CGACGCCCTG	CTGGAGCGCA	ACTATCCCAC	TGGCGCGGAG	. 1200
TTCCTCGGCG	ACGGCGGCGA	CGTCAGCTTC	AGCACCCGCG	GCAAGGGGAT	TTTAGGATTT	1260
GTGTTCACGC	TCACCGTGTG	Α				1281

# (2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 427 amino acids

(B) TYPE: amino acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

	(xi)	SEQU	JENCE	DES	SCRIF	401T¢	I: SE	Q IC	NO-:	15:						
5	Het 1	Ala	G1 u	Glu	Ala 5	Phe	Asp	Leu	Trp	Asn 10	G1 u	Cys	Ala	Lys	A1a 15	Cys
	Val	Leu	Asp	Leu 20	Lys	Asp	Gly	Val	Arg 25	Ser	Ser	Arg	Met	Ser 30	Val	Asp
. 10	Pro	Ala	Ile 35	Ala	Asp	Thr	Asn	G1 y 40	Gln	Gly	Val	Leu	His 45	Tyr	Ser	Het
	Val	Leu 50	G1 u	Gly	G1 y	Asn	Asp 55	Ala	Leu	Lys	Leu	A1a 60	Ile	Asp	Asn	Αĺα
15	Leu 65	Ser	Ile	Thr	Ser	Asp 70	Gly	Leu	Thr		Arg 75	Leu	Glu	Gly	Gly	Va1 80
20	G1u	Pro	Asn	Lys	Pro 85	Val	Arg	Tyr	Ser	Tyr 90	Thr	Arg	Gln	Ala	Arg 95	G1 y
	Ser	Trp	Ser	Leu 100	Asn	Trp	Leu	Val	Pro 105	Ile	Gly	His	Glu	Lys 110	Pro	Ser
25	Asn	Ile	Lys 115	Val	Phe	Ile	His	G1υ 120	Leu	Asn	Ala	Gly	Asn 125	Gln	Leu	Ser
	His	Met 130	Ser	Pro	Ile	Tyr	Thr 135	Ile	Glu	Met	G1 y	Asp 140	G1 u	Leu	Leu	Ala
30	Lys 145	Leu	Ala	Arg	Asp	A1a 150	Thr	Phe	Phe	Val	Arg 155	Ala	His	Glu	Ser	Asn 160
35	G1υ	Met	G1 n	Pro	Thr 165	Lev	Ala	Ile	Ser	His 170	Ala	G1 y	Val	Ser	Va1 175	Val
	Met	Ala	Gin	Thr 180	Gln	Pro	Arg	Arg	G1 u 185	Lys	Arg	Trp	Ser	G1u 190	Trp	Ala
40	Ser	Gly	Lys 195	Val	Leu	Cys	Leu	Leu 200	Asp	Pro	Leu	Asp	G1 y 205	Val	Tyr	Asn
	Tyr	Leu 210	Ala	G1n	Gln	Arg	Cys 215		Lev	Asp	Asp	Thr 220		G1u	G1 y	Lys
45	Ile 225	Tyr	Arg	Val	Leu	A1a 230		Asn	Pro	Ala	Lys 235	His	Asp	Leu	Asp	11e 240
50	Lys	Pro	Thr	Va1	11e 245		His	Arg	Leu	His 250	Phe	Pro	G1 u	Gly	G1 y 255	Ser

	•	Leu	Ala	Ala	Leu 260	Thr	Ala	His	Gln	A1a 265	Cys	Hi s	Leu	Pro	Leu 270	Glu	Thr	·
5		Phe	Thr	Arg 275	His	Arg	Gln	Pro	Arg 280	G1 y	Trp	G1 u	G1n	Leu 285	Glu	Gln	Cys	
10		Gly	Tyr 290	Pro	Val	Gln	Arg	Leu 295	Val	Ala	Leu	Tyr	Leu 300	Ala	Ala	Arg	Leu	
		Ser 305	Trp	Asn	Gln	Val	Asp 310	Gln	Val	Ile	Arg	Asn 315	Ala	Leu	Ala	Ser	Pro 320	
15		Gly	Ser	G1 y		325	Lev	G1 y	GΊυ	Ala	11e 330	Arg	G1 u	G1 n	Pro	G1 u 335	G1n	
		Ala	Arg	Leu		Leu	Thr	Leu	Ala	Ala 345	Ala	G1 u	Ser	Glu	Arg 350	Phe	Val	
20		Arg	G1 n	G1 y 355	Thr	G1 y	Asn	Asp	G1 u 360	Ala	Gly	Ala	Ala	Asn 365	Ala	Asp	Val	
25		Val	Ser 370	Leu	Thr	Cys	Pro	Va1 375	Ala	Ala	61 y	Glu	Cys 380	Ala	G1 y	Pro	Ala .	
23	•	Asp 385	Ser	Gly	Asp	Ala	Leu 390	Leu	G1 u	Arg	Asn	Tyr 395	Pro	Thr	G1 y	Ala	G1 u 400	
30		Phe	Leu	G1 y	Asp	G1 y 405	G1 y	Asp	Val	Ser	Phe 410	Ser	Thr	Arg	Gly	Lys 415	Gly	•
			Leu	G1 y	Phe 420	Val	Phe	Thr	Leu	Thr 425	Val	Xaa						
35	(2)	INFO	RMAT.	ION I	FOR :	SEQ	ID N	0:16	:									
		(i)	(A	) LEI	NGTH	: 18	TERI: basi	e pa										
40			•				SS: : line	_	1e									
		(ii)	MOL	ECULI	E TY	PE: (	DNA (	(gen	omic	)								
45																		
		(xi)				SCRI	PTIO!	N: S1	EQ II	0 МО	:16:							
50	GGCT	GATA	AT A	SAGC	TCG													18

# (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1245 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

60	GCTCGACCTC	AAGCCTGCGT -	GAATGCGCCA	CCTCTGGAAC	AAGCTTTCGA	ATGGCCGAGG
120	CACCAACGGC	CCATCGCCGA	GTCGACCCGG	CCGCATGAGC	TGCGTTCCAG	AAGGACGGCG
180	CAAGCTGGCC	ACGACGCGCT	GAGGGCGGCA	CATGGTCCTG	TGCACTACTC	CAGGGCGTGC
240	AGGCGGCGTC	TCCGCCTCGA	GGCCTGACCA	CACCAGCGAC	CCCTCAGCAT	ATCGACAACG
300	TTGGTCGCTG	CGCGCGGCAG	ACGCGCCAGG	CTACAGCTAC	AGCCGGTGCG	GAGCCGAACA
360	CATCCACGAA	TCAAGGTGTT	CCCTCGAACA	CCACGAGAAG	TACCGATCGG	AACTGGCTGG
420	GATGGGCGAC	ACACCATCGA	TCGCCGATCT	CAGCCACATG	GCAACCAGCT	CTGAACGCCG
480	CGAGAGCAAC	TCAGGGCGCA	ACCTTCTTCG	GCGCGATGCC	CGAAGCTGGC	GAGTTGCTGG
540	GGCCCAGACC	GCGTGGTCAT	GCCGGGGTCA	CATCAGCCAT	CGACGCTCGC	GAGATGCAGC
600	GTGCCTGCTC	GCAAGGTGTT	TGGGCCAGCG	CTGGAGCGAA	GGGAAAAGCG	CAGCCGCGCC
660	CGACGATACC	GCTGCAACCT	GCCCAGCAAC	CAACTACCTC	ACGGGGTCTA	GACCCGCTGG
720	CCTGGACATC	CGAAGCATGA	GGCAACCCGG	GGTGCTCGCC	AGATCTACCG	TGGGAAGGCA
780	GGCCGCGCTG	GCGGCAGCCT	TTTCCCGAGG	TCGCCTGCAC	TCATCAGTCA	AAACCCACGG
840	CCAGCCGCGC	CCCGTCATCG	GAGACTITCA	CCTGCCGCTG	AGGCTTGCCA	ACCGCGCACC
900	CCTCTACCTG	GGCTGGTCGC	CCGGTGCAGC	GTGCGGCTAT	AACTGGAGCA	GGCTGGGAAC
960	GGCCAGCCCC	GCAACGCCCT	CAGGTGATCC	CCAGGTCGAC	TGTCGTGGAA	GCGGCGCGGC
1020	CCGTCTGGCC	CGGAGCAGGC	CGCGAGCAGC	CGAAGCGATC	: GCGACCTGGG	GGCAGCGGCG

50

5

10

15

20

25

30

35

40

45

	CTGACCCTG	e ccecce	CCGA GAG	CGAGCGC	TTCGTCCG	GC AGGGC	ACCGG CA	ACGACGAG	1080
5	GCCGGCGCG	G CCAACG	CCGA CG1	GGTGAGC	CTGACCTG	CC CGGTC	cccc cc	GTGAATGC	1140
	GCGGGCCCG	G CGGACA	GCGG CGA	ACGCCCTG	CTGGAGCG	CA ACTAT	CCCAC TG	GCGCGGAG	1200
	TTCCTCGGC	G ACGGCG	GCGA CG1	CAGCTTC	AGCACCCG	CG GCTGA			1245
10	(2) INFOR	MATION F	OR SEQ	ID NO:18	:	•			
15		(A) LEN (B) TYP (C) STP (D) TOP	CHARACT IGTH: 41! PE: amino RANDEDNE! POLOGY:	amino acid SS: sing linear	acids	•.•			
20	(ii)	MOLECULE	TYPE:	protein					
20	(xi)	SEQUENCE	DESCRI	PTION: S	EQ ID NO:	: 18 :			
25					Leu Trp		Cys Ala	Lys Ala 15	Cys .
	Val	Leu Asp	Leu Lys 20	Asp Gly	Val Arg 25	Ser Ser	Arg Met	Ser Val 30	Asp
30	Pro	Ala Ile 35	Ala Asp	Thr Asn	Gly Gln 40	Gly Val	Leu His 45	Tyr Ser	Met
35	Val	Leu Glu 50	Gly Gly	Asn Asp 55	Ala Leu	Lys leu	Ala Ile 60	Asp Asn	Ala
	65			70	Leu Thr	75			80
40			85		ı Tyr Ser	90		95	
	Ser	Trp Ser	Leu Asn 100	Trp Lei	Val Pro 105		His Glu	Lys Pro 110	Ser
45	Asn	Ile Lys 115		Ile His	120 Leu	Asn Ala	Gly Asn 125	Gln Lev	Ser
	His	Met Ser 130	Pro Ile	Tyr Thi	r Ile Glu S	Met Gly	Asp Glu 140	Leu Leu	Ala
50		•							

	Lys 145	Leu	Ala	Arg	Asp	A1a 150	Thr	Phe	Phe	Val	Arg 155	Ala	His	Glu	Ser	160
5	G1 u	Het	G1n	Pro	Thr 165	Leu	Ala	Ile	Ser	Hi s 170	Ala	G1 y	Val	Ser	Va1 175	Val
	Met	Ala	Gln	Thr 180	Gln	Pro	Arg	Arg	G10 185	Lys	Arg	Trp	Ser	G1 u 190	Trp	Ala
10	Ser	G1 y	Lys 195	Val	Leu	Cys	Leu	Leu 200	Asp	Pro	Lev	Asp	G1 y 205	Val	Tyr	Asn
15	Tyr	Leu 210		Gln	Gln	Arg	Cys 215	Asn	Leu	Asp	Asp	Thr 220	Trp	G1 u	G1 y	Lys
	11 e 225		Arg	Val	Leu	A1a 230	G1 y	Asn	Pro	Alā	lys 235	His	Asp	Leu	Asp	Ile 240
20	Lys	Pro	Thr	Val	11e 245		His	Arg	Leu	His 250	Phe	Pro	Glu	G1 y	G1 y 255	Ser
	Leu	Ala	Ala	Leu 260		Ala	His	Gln	A1a 265	Cys	His	Lev	Pro	Leu 270	Glu	Thr
25	Phe	. Thr	- Arg 275		Arg	Gln	Pro	280		, Trp	G1 u	Gln	Leu 285	Glu	Gin	Cys
30	G1)	7 Tyi 290		Val	Glr	n Arg	Leu 295		A1a	s Leu	Tyr	300	Ala	Ala	Arg	, Leu
	Se:		p Asr	n G1r	n Val	310		n Val	111	e Arg	315		. Leu	Ala	Sei	Pro 320
35	G1	y Se	r G1;	y G1:	y Ası 32!		ن G1 ر	y G1	u Ala	a Ile 330	e Arg	g G1 c	, G1m	Pro	33!	, G1n 5
	Al	a Ar	g Le	u A1a 34		u Thi	r Le	u A1	a A1 34		a Glu	) Sei	r Glu	350	g Pho	e Val
40	Ar	g G1	n G1 35	_	r G1	y As	n As	p G1 36	u A1 0	a Gl	у А1;	a Ala	a Asr 365	n A1a 5	a As	p Val
45	۷a	1 Se		u Th	r Cy	s Pr	o Va 37		a Al	a G1	y G1	u Cy 38	s A1a O	a G1;	y Pr	o Ala
7-7	As 38	_	er G1	y As	p Al	a Le 39		u G1	u Ar	g As	n Ty 39	r Pr 5	o Th	r G1	y Al	a Glu 400

	Phe Leu Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Xaa 405 410 415	
5	(2) INFORMATION FOR SEQ ID NO:19:	
10	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 25 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
	(ii) MOLECULE TYPE: DNA (genomic)	
15		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:	
	TCGAGCCGCC ACCATGGCCG AGGAA	25
20	(2) INFORMATION FOR SEQ ID NO:20:	
25	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 46 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
	(ii) MOLECULE TYPE: DNA (genomic)	
30		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:	
35	GACCCGCTAG CACCCGGGAA ACCGCCGCGC GAGGACCTGA ÁGTAAG	46
	(2) INFORMATION FOR SEQ ID NO:21:	
40	<ul> <li>(i) SEQUENCE CHARACTERISTICS:         <ul> <li>(A) LENGTH: 1956 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul> </li> </ul>	
45	(ii) MOLECULE TYPE: DNA (genomic)	
50		

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

	ATGCACCTGA	TACCCCATTG	GATCCCCCTG	GTCGCCAGCC	TCGGCCTGCT	CGCCGGCGGC	60
5	TCGTCCGCGT	CCGCCGCCGA	GGAAGCTTTC	GACCTCTGGA	ACGAATGCGC	CAAAGCCTGC	120
	GTGCTCGACC	TCAAGGACGG	CGTGCGTTCC	AGCCGCATGA	GCGTCGACCC	GGCCATCGCC	180
10	GACACCAACG	GCCAGGGCGT	GCTGCACTAC	TCCATGGTCC	TGGAGGGCGG	CAACGACGCG	240
	CTCAAGCTGG	CCATCGACAA	CGCCCTCAGC	ATCACCAGCG	ACGGCCTGAC	CATCCGCCTC	300
	GAAGGCGGCG	TCGAGCCGAA	CAAGCCGGTG	CGCTACAGCT	ACACGCGCCA	GGCGCGCGGC	360
15	AGTTGGTCGC	TGAACTGGCT	GGTACCGATC	GGCCACGAGA	AGCCCTCGAA	CATCAAGGTG	420
	TTCATCCACG	AACTGAACGC	CGGCAACCAG	CTCAGCCACA	TGTCGCCGAT	CTACACCATC	480
20	GAGATGGGCG	ACGAGTTGCT	GGCGAAGCTG	GCGCGCGATG	CCACCTTCTT	CGTCAGGGCG	540
	CACGAGAGCA	ACGAGATGCA	GCCGACGCTC	GCCATCAGCC	ATGCCGGGGT	CAGCGTGGTC	600
	ATGGCCCAGA	CCCAGCCGCG	CCGGGAAAAG	CGCTGGAGCG	AATGGGCCAG	CGGCAAGGTG	660
25	TTGTGCCTGC	TCGACCCGCT	GGACGGGGTC	TACAACTACC	TCGCCCAGCA	ACGCTGCAAC	720
	CTCGACGATA	CCT.GGGAAGG	CAAGATCTAC	CGGGTGCTCG	CCGGCAACCC	GGCGAAGCAT	780
30	GACCTGGACA	TCAAACCCAC	GGTCATCAGT	CATCGCCTGC	ACTTTCCCGA	GGGCGGCAGC	840
	CTGGCCGCGC	TGACCGCGCA	CCAGGCTTGC	ÇACCTGCCGC	TGGAGACTTT	CACCCGTCAT	900
	CGCCAGCCGC	GCGGCTGGGA	ACAACTGGAG	CAGTGCGGCT	ATCCGGTGCA	GCGGCTGGTC	960
35	GCCCTCTACC	TGGCGGCGCG	GCTGTCGTGG	AACCAGGTCG	ACCAGGTGAT	CCGCAACGCC	1020
	CTGGCCAGCC	CCGGCAGCGG	CGGCGACCTG	GGCGAAGCGA	TCCGCGAGCA	GCCGGAGCAG	1080
40	GCCCGTCTGG	CCCTGACCCT	GGCCGCCGCC	GAGAGCGAGC	GCTTCGTCCG	GCAGGGCACC	1140
40	GGCAACGACG	AGGCCGGCGC	GGCCAACGCC	GACGTGGTGA	GCCTGACCTG	CCCGGTCGCC	1200
	GCCGGTGAAT	GCGCGGGCCC	GGCGGACAGC	GGCGACGCCC	TGCTGGAGCG	CAACTATCCC	126
45	ACTGGCGCGG	AGTTCCTCGG	CGACGGCGGC	GACGTCAGCT	TCAGCACCCG	CGGCACGCAG	1326
	AACTGGACGG	TGGAGCGGCT	GCTCCAGGCG	CACCGCCAAC	TGGAGGAGCG	CGGCTATGTG	138

50

	TICGTCGGCT ACCACGGCAC CTTCCTCGAA GCGGCGCAAA GCATCGTCTT CGGCGGGGTG	1440
5	CGCGCGCGCA GCCAGGACCT CGACGCGATC TGGCGCGGTT TCTATATCGC CGGCGATCCG	1500
	GCGCTGGCCT ACGGCTACGC CCAGGACCAG GAACCCGACG CACGCGGCCG GATCCGCAAC	1560
	GGTGCCCTGC TGCGGGTCTA TGTGCCGCGC TCGAGCCTGC CGGGCTTCTA CCGCACCAGC	1620
10	CTGACCCTGG CCGCGCCGGA GGCGGCGGGC GAGGTCGAAC GGCTGATCGG CCATCCGCTG	1680
	CCGCTGCGCC TGGACGCCAT CACCGGCCCC GAGGAGGAAG GCGGGCGCCT GGAGACCATT	1740
15	CTCGGCTGGC CGCTGGCCGA GCGCACCGTG GTGATTCCCT CGGCGATCCC CACCGACCCG	1800
	CGCAACGTCG GCGGCGACCT CGACCCGTCC AGCATCCCCG ACAAGGAACA GGCGATCAGC	1860
	GCCCTGCCGG ACTACGCCAG CCAGCCCGGC AAACCGCCGC GCGAGGACCC GCTAGCACCC	1920
20	GGGAAACCGC CGCGCGAGGA CCTGAAGTAA GAATTC	1956
	(2) INFORMATION FOR SEQ ID NO:22:	
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 652 amino acids	
25	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30	(ii) MOLECULE TYPE: protein	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
35	Met His Leu Ile Pro His Trp Ile Pro Leu Val Ala Ser Leu Gly Leu	
	1 5 10 15	
	Leu Ala Gly Gly Ser Ser Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu 20 25 30	
40		
	Trp Asn Glu Cys Ala Lys Ala Cys Val Leu Asp Leu Lys Asp Gly Val 35 40 45	
45	Arg Ser Ser Arg Met Ser Val Asp Pro Ala Ile Ala Asp Thr Asn Gly 50 55 60	
	Gln Gly Val Leu His Tyr Ser Met Val Leu Glu Gly Gly Asn Asp Ala 65 70 75 80	
50		

	Leu	Lys	Leu		I1e 85	Asp	Asn	Ala	Lev	Ser 90	Ile	Thr	Ser	Asp	G1 y 95	Leu
5	Thr	Ile	Arg	Leu 100	Glυ	Gly	G1 y	Val	G1 u 105	Pro	Asn	Lys	Pro	Va1 110	Arg	Tyr
	Ser	Tyr	Thr 115	Arg	Gln	Ala		G1 y 120	Ser	T-rp	Ser	Leu	Asn 125	Trp	Leu	Val
10	Pro	I1e 130	G1 y	His	Glu	Lys	Pro 135	Ser	Asn	Ile	Lys	Val 140	Phe	Ile	His	Glu
15	Leu 145	Asn	Αla	Gly	Aşn	G1n 150	Leu	Ser	His	Met	Ser 155	Pro	Ile	Tyr	Thr	Ile 160
	Glu	Met	Gly	Asp	G1 u 165	Leu	Leu	Αla	Lys	Leu 170	s FA	Arg	Asp	Ala	Thr 175	Phe
20	Phe	Val	Arg	A1a 180	Hi <sub>5</sub>	Glu	Ser	Asn	G1 u 185	Met	Gln	·Pro	Thr	Leu 190	Ala	Ile
	Ser	His	A1 a		۷a۱	Ser	Val	Va1 200		Ala	Gln	Thr	G1n 205	Pro	Arg	Arg
25	Glu	Lys 210		Trp	Ser	Glu	Trp 215	Ala	Ser	· G1 y	Lys	Va1 220	Leu	Cys	Leu	Leu
20	Asp 225		Leu	Asp	Gly	Va1 230		Asn	Tyr	· Leu	A1a 235	G1 n	G1n	Arg	Cys	Asn 240
30	Leu	Asp	Asp	Thr	7rp 245		Gly	Lys	: 11	250	Arg	Val	Leu	ı Ala	G1 <sub>3</sub> 255	Asn
35	Pro	) A1a	a Lys	260		Leu	. Asp	Ιlε	26!		Thr	· Val	<b>  11</b> 6	270	. Hi:	; Arg
	Leu	Hi:	s Pho 27!		G1 (	رG۱ د	/ Gly	Se (		u Ala	Ala	Le	285	- Ala	. Hi:	s Gln
40	Ala	a Cy:		s Le	u Pri	o Lei	ن G1 د 295		r Ph	e Thi	- Arg	30(		g G1r	ı Pr	Arg
	G1 :		p G1	u G1	n Le	u G1:		Cy	s G1	у Туі	7 Pro		1 G1:	n Arg	j Le	v Val 320
45	<b>A</b> 1:	a le	υТу	r le	u A1 32		a Arg	g Le	υ Se	r Tr <sub>i</sub> 330		n G1	n Va	1 Ası	9 G1 33	n Val 5
50																

	Ile	Arg	Asn	A1a 340	Lev	Ala	Ser	Pro	G1 y 345	Ser	G1 y	G1 y	Asp	Leu 350	G1 y	G1 u
5	Ala	Ile	Arg 355	G1 u	Gln	Pro	Glυ	G1n 360	Ala	Arg	Leu	Ala	Leu 365	Thr	Leu	Ala
10	Ala	A1a 370	Glu	Ser	Glu	Arg	Phe 375	Val	Arg	Gln	G1 y	Thr 380	G1 y	Asn	Asp	Glu
	A1a 385	G1 y	Ala	Ala	Asn	Ala 390		۷a۱	Val	Ser	Leu 395	Thr	Cys	Pro	Val	A1a 400
. 15	Ala	G1 y	Glu	Cys	405	G1 y	Pro	Ala	Asp	Ser 410	G1 y	Asp	Ala	Leu	Leu 415	Glu
	Arg	Asn	Tyr			Gly	Ala	<b>G1</b> u	Phe 425	Leu	G1 y	Asp	Gly	G1 y 430	Asp	۷a۱
20	Ser	Phe	Ser 435	Thr	Arg	Gly	Thr	G1n 440	Asn	Trp	Thr	Va1	G1u 445	Arg	Lev	Lev
25	Gln	A1a 450	His	Arg	G1 n	Leu	G1ս 455	Glu	Arg	Gly	Tyr	Va1 460	Phe	Va1	G1 y	Tyr
20	His 465		Thr	Phe	Leu	G1υ 470	Ala	Ala	G1 n	Ser	Ile 475	Val	Phe	G1 y	G1 y	Va1 480
30	Arg	Ala	Arg	Ser	G1 n 485	Asp	Lev	Asp	Αla	Ile 490	Trp	Arg	Gly	Phe	Tyr 495	Ile
	Ala	G1 y	Asp	Pro 500	Ala	Leu	Ala	Tyr	G1 y 505		Ala	Gln	Asp	G1n 510	Glu	Pro
35	Asp	Δla	Arg 515		Arg	Ile	Arg	Asn 520		Ala	Leu	Leu	Arg 525		Tyr	Val
	Pro	Arg 530		Ser	Leu	Pro	G1 y 535		Tyr	Arg	Thr	Ser 540		Thr	Leu	Ala
40	A1a 545		Glu	Ala	Ala	G1 y 550		Val	Glu	Arg	Leu 555		G1 y	His	Pro	Leu 560
45	Pro	Leu	Arg	Leu	Asp 565		Ile	Thr	G1 y	Pro 570		G1 u	G1u	G1 y	G1 y 575	Arg
	Leu	G1u	Thr	I1e 580		Gly	Trp	Pro	Leu 585		Glu	Arg	, Thr	Va1 590		Ile
50																

	Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp 595 600 605	
5	Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp 610 615 620	
	Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Pro Leu Ala Pro 625 630 635 640	
10	Gly Lys Pro Pro Arg Glu Asp Leu Lys Xaa Glu Phe 645 650	
	(2) INFORMATION FOR SEQ ID NO:23:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 48 base pairs  (B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
	CCGGGCTGAC TAAGGGGATT TTAGGATTTG TGTTCACGCT CACCGTGC	48
30	(2) INFORMATION FOR SEQ ID NO:24:	
	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 2004 base pairs</li><li>(B) TYPE: nucleic acid</li></ul>	
35	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
	ATGCACCTGA TACCCCATTG GATCCCCCTG GTCGCCAGCC TCGGCCTGCT CGCCGGCGGC	60
45	TCGTCCGCGT CCGCCGCCGA GGAAGCTTTC GACCTCTGGA ACGAATGCGC CAAAGCCTGC	120
	GTGCTCGACC TCAAGGACGG CGTGCGTTCC AGCCGCATGA GCGTCGACCC GGCCATCGCC	180
50		

	GACACCAACG GCCAGGGCGT GCTGCACTAC TCCATGGTCC TGGAGGGCGG CAACGACGCG	240
	CTCAAGCTGG CCATCGACAA CGCCCTCAGC ATCACCAGCG ACGGCCTGAC CATCCGCCTC	300
5	GAAGGCGGCG TCGAGCCGAA CAAGCCGGTG CGCTACAGCT ACACGCGCCA GGCGCGCGC	360
	AGTTGGTCGC TGAACTGGCT GGTACCGATC GGCCACGAGA AGCCCTCGAA CATCAAGGTG	420
10	TTCATCCACG AACTGAACGC CGGCAACCAG CTCAGCCACA TGTCGCCGAT CTACACCATC	480
	GAGATGGGCG ACGAGTTGCT GGCGAAGCTG GCGCGCGATG CCACCTTCTT CGTCAGGGCG	540
	CACGAGAGCA ACGAGATGCA GCCGACGCTC GCCATCAGCC ATGCCGGGGT CAGCGTGGTC	600
15	ATGGCCCAGA CCCAGCCGCG CCGGGAAAAG CGCTGGAGCG AATGGGCCAG CGGCAAGGTG	660
	TIGTGCCTGC TCGACCCGCT GGACGGGGTC TACAACTACC TCGCCCAGCA ACGCTGCAAC	720
20	CTCGACGATA CCTGGGAAGG CAAGATCTAC CGGGTGCTCG CCGGCAACCC GGCGAAGCAT	780
20	GACCTGGACA TCAAACCCAC GGTCATCAGT CATCGCCTGC ACTTTCCCGA GGGCGGCAGC	840
	CTGGCCGCGC TGACCGCGCA CCAGGCTTGC CACCTGCCGC TGGAGACTTT CACCCGTCAT	900
25	CGCCAGCCGC GCGGCTGGGA ACAACTGGAG CAGTGCGGCT ATCCGGTGCA GCGGCTGGTC	960
	GCCCTCTACC TGGCGGCGCG GCTGTCGTGG AACCAGGTCG ACCAGGTGAT CCGCAACGCC	1020
20	CTGGCCAGCC CCGGCAGCGG CGGCGACCTG GGCGAAGCGA TCCGCGAGCA GCCGGAGCAG	1080
30	GCCCGTCTGG CCCTGACCCT GGCCGCCGCC GAGAGCGAGC GCTTCGTCCG GCAGGGCACC	1140
	GGCAACGACG AGGCCGGCGC GGCCAACGCC GACGTGGTGA ÇCCTGACCTG CCCGGTCGCC	1200
35	GCCGGTGAAT GCGCGGGCCC GGCGGACAGC GGCGACGCCC TGCTGGAGCG CAACTATCCC	1260
	ACTGGCGCGG AGTTCCTCGG CGACGGCGGC GACGTCAGCT TCAGCACCCG CGGCACGCAG	1320
	AACTGGACGG TGGAGCGGCT GCTCCAGGCG CACCGCCAAC TGGAGGAGCG CGGCTATGTG	1380
40	TTCGTCGGCT ACCACGGCAC CTTCCTCGAA GCGGCGCAAA GCATCGTCTT CGGCGGGGTG	1440
	CGCGCGCGCA GCCAGGACCI CGACGCGATC TGGCGCGGTT TCTATATCGC CGGCGATCCG	1500
45	GCGCTGGCCT ACGGCTACGC CCAGGACCAG GAACCCGACG CACGCGGCCG GATCCGCAAC	1560
	GGIGCCCIGC IGCGGGICTA IGTGCCGCGC ICCAGCCTGC CGGGCTTCTA CCGCACCAGC	1620

	CTGACCC	rgg co	cgcgc	CGGA	GGC	GGCG	iGGC	GAGG	TCGA	AC 6	GCTG	ATCG	G CC	ATCC	GCTG	i	1680
_	CCGCTGC	GCC TO	GGACG	CCAT	CAC	CGGC	ccc	GAGG	AGGA	AG G	cece	CGCC	T GG	AGAC	CATT		1740
5	CTCGGCT	GGC C	GCTGG	CCGA	GCG	CACC	GTG	GTGA	TTCC	ст с	GGCG	ATCC	C CA	CCGA	cccg	i	1800
	CGCAACG	TCG G	cGGCG	ACCT	CGA	ccce	TCC	AGCA	тссс	CG A	CAAG	GAAC	A GG	CGAT	CAGC	:	1860
10	ĠCCCTGC	CGG A	CTAC	CCAG	CCA	GCCC	GGC	AAAC	CGCC	GC (	CGAG	GACC	c GC	TAGO	ACCC		1920
	GGGCTGA	CTA A	GGGGA	TTTT	AGG	ATTI	GTG	TTCA	CGCT	CA (	CGTG	cccg	G GA	AACC	GCCG	;	1980
	CGCGAGG	ACC T	GAAGT	AAGA	ATT	c											2004
15	(2) INF	ORMAT	ION F	OR	EQ I	D NO	):25:	:			•						
20		(B (C (D	) LEN ) TYP ) STR ) TOP	IGTH: PE: a RANDE POLOG	668 mino DNES	ami aci iS: s inea	ino a id singl	acids	į								
25	(ii	) MOL	ECULE	TYP	'E: p	rote	ein										
	(xi	) SEQ	UENC	DES	CRIF	PT I OI	4: SI	EQ 16	) NO:	25:							
30	Me 1	t His	Leu	Ile	Pro 5	His	Trp	IJе	Pro	Leu 10	Val	Ala	Ser	Lev	G1 y 15	Leu	
	Le	u Ala	Gly	G1 y 20	Ser	Ser	Ala	Ser	A1a 25	A1a	Glυ	G1 u	Ala	Phe 30	Asp	Leu	
35	Tr	p Asn	61 ս 35	Cys	Ala	Lys	Ala	Cys 40	Val	Leu	Asp	Leu	Lys 45	Asp	G1 y	Val	
40	Ar	g Ser 50	Ser	Arg	Met	Ser	Va 1 55	Asp	Pro .	Ala	Ile	A1a 60	Asp	Thr	Asn	G1 y	
	G1 65	n Gly	Va1	Leu	His	Tyr 70	Ser	Met	Va1	Leu	G1υ 75	G1 y	G1 y	Asn	Asp	A1a 80	
45	Le	u Lys	Leu	Ala	I1e 85	Asp	Asn	Ala	Leu	Ser 90	Ile	Thr	Ser	Asp	G1 y 95	Leu	
	Th	r Ile	Arg	Leu 100	Glu	Gly	G1 y	Val	G1 u 1 <b>0</b> 5	Pro	Asn	Lys	Pro	Va1 110	Arg	Tyr	
50																	

	Ser	Tyr	115		G1n	Ala	Arg	120		Trp	Ser	Leu	Asn 125		Leu	Va1
<i>5</i>	Pro	11e	_	His	Glu	Lys	Pro 135		Asn	Ile	Lys	Va1 140	Phe	Ile	His	G1u
10	Leu 145		Ala	Gly	Asn	G1n 150		Ser	His	Het	Ser 155		Ile	Tyr	Thr	Ile 160
	Glu	Het	G1 y	Asp	G1u 165	Leu	Leu	Ala	Lys	Leu 170		Arg	Asp	Ala	Thr 175	Phe
15	Phe	Va1	Arg	A1a 180		G1υ	Ser	Asn	G1 u 185		G1n	Pro	Thr	Leu 190	Ala	Ile
	Ser	His	Ala 195	Gly	۷a۱	Ser	Val	Va1 200	Met	Ala	Gln	Thr	G1 n 205	Pro	Arg	Arg
20	Glu	Lys 210	Arg	Trp	Ser	Glu	Trp 215		Ser	Gly	Lys	Va1 220	Lev	Cys	Leu	Leu
25	<b>Asp</b> 225	Pro	Leu	Asp	G1 y	Va1 230	Tyr	Asn	Tyr	Leu	Ala 235	Gln	Gln	Arg	Cys	Asn 240
	Leu	Asp	Asp	Thr	Trp 245	Glu	Gly	Lys	Ile	Tyr 250	Arg	Val	Leu	Ala	G1 y 255	Asn
30	Pro	Ala	Lys	His 260	Asp	Leu	Asp	Ile	Lys 265	Pro	Thr	Val	Ile	Ser 270	His	Arg
	Leu	Hi s	Phe 275	Pro	Glυ	Gly	G1 y	Ser 280	Leu	Ala	Ala	Leu	Thr 285	Ala	His	G1n
35	Ala	Cys 290	His	Lev	Pro	Leu	G1 u 295	Thr	Phe	Thr	Arg	His 300	Arg	G1n	Pro	Arg
	G1 y 305	Тгр	GΊυ	G1n	Leu	G1u 310	Gin	Cys	Gly	Tyr	Pro 315	Val	G1 n	Arg	Leu	Va1 320
40	Ala	Lev	Tyr	Leu	A1a 325	Ala	Arg	Leu	Ser	Trp 330	Asn	Gln	Val	Asp	G1n 335	Val
45	Ile	Arg	Asn	A1a 340	Leu	Ala	Ser	Pro	G1 y 345	Ser	G1 y	Gly	Asp	Leu 350	Gly	G1 u
	Ala:	Ile	Arg 355	Glu	G1 n	Pro	G1 u	G1 n 360	Ala	Arg	Leu		Leu 365	Thr	Leu	Ala
50																

	Ala	37(	a Glu O	, Ser	· G1	) Ar	g Pho 379		l Arq	g G1n	G1,	7 Thi 380	_	y Asr	n Asp	G1u
5	A1 a		y Ala	ı Ala	Asr	390		Val	l Val	l Ser	1 Leu		- Cys	Pro	Val	Ala 400
•	Αla	( G1 y	y Glu	Cys	405		, Pro	Ala	a Asp	5- Ser 410		Asp	Ala	. Leu	Leu 415	
10	Arg	Aşr	ı Tyr	Pro 420		· 61,	/ Ala	( G1 t	9 Phe 425		G1 y	Asp	Gly	G1 y 430		Val
15	Ser	Phe	9 Ser 435		Arg	G1 y	Thr	G1 n		Trp	Thr	Val	G1 u 445	-		Leu '
	Gln	A1 a 450	His	·Arg	Gln	Leu	G1u 455		Arg	G1 y	Tyr	Va1 460		Val	G1 y	Tyr
20	His 465	G1 y	Thr	Phe	Leu	G1 u 470		Αla	Gln	Ser	I1e 475	Va1	Phe	G1 y	G1 y	Va1 480
	Arg	Ala	Arg	Ser	G1 n 485		Lev	Asp	Ala	Ile 490	Trp	Arg	G1 y	Phe	Tyr 495	Ile
25	Αlą	Gly	Asp	Pro 500	Ala	Leu	Ala	Tyr	G1 y 505	Tyr	Ala	Gln	Asp	G1n 510	Glu	Pro
<i>30</i>	Asp	Ala	Arg 515	G1 y	Arg	Ile	Arg	Asn 520	G1 y	Ala	Leu	Leu	Arg 525	Val	Tyr	Val
	Pro	Arg 530	Ser	Ser	Leu	Pro	G1 y 535	Phe	Tyr	Arg	Thr	Ser 540	Lev	Thr	Leu	Ala
35	A1a 545	Pro	Glu	Ala	Ala	Gly 550	Glu	Va1	GΊυ	Arg	Leu 555	Ile	Gly	Hi s	Pro	Leu 560
	Pro	Lev	Arg	Lev	<b>Asp</b> 565	Ala	Ile	Thr	G1 y	Pro 570	G1 u	G1 u	G1 u	Gly	G1 y 575	Arg
40	Leu	Glu	Thr	I1e 580	Leu	Gly	Trp	Pro	Leu 585	Ala	G1 u	Arg	Thr	Va1 590	Val	Ile
_	Pro	Ser	A1a 595	Ile	Pro	Thr	Asp	Pro 600	Arg	Asn	Val	Gly	G1 y 605	Asp	Leu	Asp
45		Ser 610	Ser	Ile	Pro	Asp	Lys 615	G1 u	Gln	Ala		Ser 620	Ala	Leu	Pro	Asp
50																

	Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Pro Leu Ala Pro 625 630 636 635 640	
5	Gly Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr Val Pro 645 650 655	
	Gly Lys Pro Pro Arg Glu Asp Leu Lys Xaa-Glu Phe 660 665	
10	(2) INFORMATION FOR SEQ ID NO:26:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 27 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
20		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
25	GCACCCGGGA TCCCGTCAGG CCCCCTC	27
	(2) INFORMATION FOR SEQ ID NO:27:	
30	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 27 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li><li>(D) TOPOLOGY: linear</li></ul>	
35	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
40	GCACCCGGGC TCCCTCTTGA GCTTCCT	27
	(2) INFORMATION FOR SEQ ID NO:28:	
45	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 2238 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
50		

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

60	cecceeceec	TCGGCCTGCT	GTCGCCAGCE	GATCCCCCTG	TACCCCATTG	ATGCACCTGA	
120	CAAAGCCTGC	ACGAATGCGC	GACCTCTGGA	GGAAGCTTTC	CCGCCGCCGA	TCGTCCGCGT	
180	GGCCATCGCC	GCGTCGACCC	AGCCGCATGA	сстсссттсс	TCAAGGACGG	GTGCTCGACC	
240	CAACGACGCG	TGGAGGGCGG	TCCATGGTCC	GCTGCACTAC	GCCAGGGCGT	GACACCAACG	
300	CATCCGCCTC	ACGGCCTGAC	ATCACCAGCG	CGCCCTCAGC	CCATCGACAA	CTCAAGCTGG	
360	eecececec	ACACGCGCCA	CGCTACAGCT	CAAGCCGGTG	TCGAGCCGAA	GAAGGCGGCG	
420	CATCAAGGTG	AGCCCTCGAA	GGCCACGAGA	GGTACCGATC	TGAACTGGCT	AGTTGGTCGC	
480	CTACACCATC	TGTCGCCGAT	CTCAGCCACA	CGGCAACCAG	AACTGAACGC	TTCATCCACG	
540	CGTCAGGGCG	CCACCTTCTT	GCGCGCGATG	GGCGAAGCTG	ACGAGTTGCT	GAGATGGGCG	
600	CAGCGTGGTC	ATGCCGGGGT	GCCATCAGCC	GCCGACGCTC	ACGAGATGCA	CACGAGAGCA	
660	CGGCAAGGTG	AATGGGCCAG	CGCTGGAGCG	CCGGGAAAAG	CCCAGCCGCG	ATGGCCCAGA	
720	ACGCTGCAAC	TCGCCCAGCA	TACAACTACC	GGACGGGGTC	TCGACCCGCT	TTGTGCCTGC	
780	GGCGAAGCAT	CCGGCAACCC	CGGGTGCTCG	CAAGATCTAC	CCTGGGAAGG	CTCGACGATA	
840	GGGCGGCAGC	ACTTTCCCGA	CATCGCCTGC	GGTCATCAGT	TCAAACCCAC	GACCTGGACA	
900	CACCCGTCAT	TGGAGACTTT	CACCTGCCGC	CCAGGCTTGC	TGACCGCGCA	CTGGCCGCGC	
960	GCGGCTGGTC	ATCCGGTGCA	CAGTGCGGCT	ACAACTGGAG	GCGGCTGGGA	CGCCAGCCGC	
1020	CCGCAACGCC	ACCAGGTGAT	AACCAGGTCG	GCTGTCGTGG	TGGCGGCGCG	GCCCTCTACC	
1080	GCCGGAGCAG	TCCGCGAGCA	GGCGAAGCGA	CGGCGACCTG	CCGGCAGCGG	CTGGCCAGCC	
1140	GCAGGGCACC	GCTTCGTCCG	GAGAGCGAGC	GCCCCCCCC	CCCTGACCCT	GCCCGTCTGG	
1200	CCCGGTCGCC	GCCTGACCTG	GACGTGGTGA	GGCCAACGCC	AGGCCGGCGC	GGCAACGACG	
1260	CAACTATCCC	TGCTGGAGCG	GGCGACGCCC	GGCGGACAGC	GCGCGGGCCC	GCCGGTGAAT	

	ACTGGCGCGG	AGTTCCTCGG	CGACGGCGGC	GACGTCAGCT	TCAGCACCCG	CGGCACGCAG	1320
5	AACTGGACGG	TGGAGCGGCT	GCTCCAGGCG	CACCGCCAAC	TGGAGGAGCG	CGGCTATGTG	1380
v	TTCGTCGGCT	ACCACGGCAC	CTTCCTCGAA	GCGGCGCAAA	GCATCGTCTT	CGGCGGGGTG	1440
	CGCGCGCGCA	GCCAGGACCT	CGACGCGATC	TGGCGCGGTT	TCTATATCGC	CGGCGATCCG	1500
10	GCGCTGGCCT	ACGGCTACGC	CCAGGACCAG	GAACCCGACG	CACGCGGCCG	GATCCGCAAC	1560
	GGTGCCCTGC	TGCGGGTCTA	TGTGCCGCGC	TCGAGCCTGC	CGGGCTTCTA	CCGCACCAGC	1620
15	CTGACCCTGG	CCGCGCCGGA	GGCGGCGGGC	GAGGTCGAAC	GGCTGATCGG	CCATCCGCTG	1680
15	CCGCTGCGCC	TGGACGCCAT	CACCGGCCCC	GAGGAGGAAG	CCGGCCCCT	GGAGACCATT	1740
	CTCGGCTGGC	CGCTGGCCGA	GCGCACCGTG	GTGATTCCCT	CGGCGATCCC	CACCGACCCG	1800
20	CGCAACGTCG	GCGGCGACCT	CGACCCGTCC	AGCATCCCCG	ACAAGGAACA	GGCGATCAGC	1860
	GCCCTGCCGG	ACTACGCCAG	CCAGCCCGGC	AAACCGCCGC	GCGAGGACCC	GCTAGCACCC	1920
	GGGATCCCGT	CAGGCCCCCT	CAAAGCCGAG	ATCGCACAGA	GACTTGAAGA	TGTCTTTGCA	1980
25	GGGAAGAACA	CCGATCTTGA	GGTTCTCATG	GAATGGCTAA	AGACAAGACC	AATCCTGTCA	2040
	CCTCTGACTA	AGGGGATTTT	AGGATTIGTG	TTCACGCTCA	CCGTGCCCAG	TGAGCGAGGA	2100
30	CTGCAGCGTA	GACGCTTTGT	CCAAAATGCC	CTTAATGGGA	ACGGGGATCC	AAATAACATG	2160
	GACAAAGCAG	TTAAACTGTA	TAGGAAGCTC	AAGAGGGAGC	CCGGGAAACC	GCCGCGCGAG	2220
	GACCTGAAGT	AAGAATTC					2238

(2) INFORMATION FOR SEQ ID NO:29:

# (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 746 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

45

35

40

50

	(xi)	SEQ	UENC	E DE	SCRI	PTIO	N: S	EQ I	D NO	:29:						
5	Met 1	His	Leu	Ile	Pro 5	His	Trp	Ile	Pro	Leu 10	Val	Ala	Ser	Leu	G1 y 15	Leu
·	Leu	Ala	Gly	G1 y 20	Ser	Ser	Αla	Ser	A1 a 25	Ala	Glu	Glu	Ala	Phe 30	Asp	Leu
10	Trp	Asn	G1 u 35	Cys	Ala	Lys	Ala	.Cys 40	Val	Leu	Asp	Leu	Lys 45	Asp	Gly	Val
15	Arg	Ser 50	Ser	Arg	Met	Ser	Va 1 55	Asp	Pro	Ala	Ile	A1a 60	Asp	Thr	Asn	G1 y
,,	G1 n 65	Gly	Val	teu	His	Tyr 70	Ser	Met	Val	Lev	GԴս 75	Gly	Gly	Asn	Asp	A1a 80
20	Leu	Lys	Lev	Ala	11e 85	Asp	Asn	Ala	Leu	Ser 90	Ile	Thr	Ser	Asp	G1 y 95	Leu
	Thr	Пe	Arg	Leu 100	Glu	G1 y	G1 y	Va1	G1 u 105	Pro	Asn	Lys	Pro	Va1 110	Arg	Tyr
25	Ser	Tyr	Thr 115	Arg	Gln	Ala	Arg	G1 y 120	Ser	Trp	Ser	Leu	Asn 125	Trp	Leu	Val
	Pro	Ile 130	G1 y	Hi s	G1 u	Lys	Pro 135	Ser	Asn	Пe	Lys	Va1 140	Phe	Ile	His	Glu
30	Leu 145	Asn	Ala	G1 y	Asn	G1n 150	Leu	Ser	His	Met	Ser 155	Pro	Ile	Tyr	Thr	Ile 160
35	Glυ	Met	Gly	Asp	G1u 165	Leu	Leu	Ala	Lys	Leu 170	Ala	Arg	Asp	Ala	Thr 175	Phe
	Phe	Val	Arg	A1a 180	His	GΊυ	Ser	Asn	G1u 185	Met	G1n	Pro	Thr	Leu 190	Ala	Ile
40	Ser	His	A1a 195	Gly	Val	Ser	Val	Va1 200	Met	Δla	G1 n	Thr	G1 n 205	Pro	Arg	Arg
	Glu	Lys 210	Arg	Trp	Ser	G1 u	Trp 215	Ala	Ser	G1 y	Lys	Va1 220	Leu	Cys	Leu	Leu
<b>4</b> 5	Asp 225	Pro	Leu	Asp	G1 y	Va 1 230	Tyr	Asn	Tyr	Lev	A1a 235	Gln	Gln	Arg	Cys	Asn 240
50	Leu	Asp	Asp	Thr	Trp 245	G1υ	Gly	Lys	Ile	Tyr 250	Arg	Val	Leu	Ala	G1 y 255	Asn

	Pr	o Al	a Ly	s Hi 26		p Le	u Asi	) I1	e Ly 26		o Th	r Va	1 11	e Se 27		s Arg
5	Le	u Hi	s Ph 27		o G1	υ G1 <u>:</u>	y Gly	y Sei 280		u Ala	a A1	a Le	u Th 28:		a Hi	s G1n
10	Al	а Су 29		s Le	u Pro	o Lei	u G1 u 295		- Pho	e Th	r Ar	9 Hi: 300		g G1	n Pr	o Arg
	G1 30	y Tr 5	p G1	u G1:	n Lei	310		Cys	G1 y	у Ту	7 Pro		l G1r	n Ar	g Le	Va1 320
15	Al	a le	υ Τγ	r lei	, A1a 325		ı Arç	. Leu	Sei	7 Trp 330		1 G1r	ı Val	l Ası	9 G1r 335	val
	11	e Ar	g Asr	340		ı Ala	Ser	Pro	G1 y 345		· G1 y	Gly	Asp	350	_	G1u
20	Ala	a Ile	355		G1n	Pro	G1u	G1 n 360	Ala	Arg	Leu	Ala	Leu 365		- Leu	Ala
25	Ala	370	1 G1u	Ser	Glu	Arg	Phe 375	Val	Arg	Gln	G1 y	Thr 380		Asn	Asp	Glu
	A1 a 385	G1 y	Ala	Ala	Asn	A1a 390	Asp	Va1	Val	Ser	Leu 395	Thr	Cys	Pro	Val	Ala 400 -
30	Αla	. G1 y	Glu	Cys	Ala 405	Gly	Pro	Ala	Asp	Ser 410	G1 y	Asp	Ala	Leu	Leu 415	G1υ
	Arg	Asn	Tyr	Pro 420	Thr	G1 y	Ala	G1'u	Phe 425	Leu	Gly	Asp	G1 y	G1 y 430	Asp	Val
35	Ser	Phe	Ser 435	Thr	Arg	Gly	Thr	G1 n 440	Asn	Trp	Thr	Val	G1 u 445	Arg	Leu	Leu
<b>4</b> 0	Gln	A1a 450	His	Arg	G1n	Leu	G1u 455	G1 u	Arg	Gly	Tyr	Va1 460	Phe	Val	Gly	Tyr
***	His 465	Gly	Thr	Phe		G1 u 470	Ala	Ala	Gin	Ser	Ile 475	Va1	Phe	G1 y	G1 y	Va1 480
45	Arg	Ala	Arg	Ser	G1 n 485	Asp	Leu .	Asp		Ile 490	Trp	Arg	G1 y	Phe	Tyr 495	Ile
	Ala	Gly	Asp	Pro 500	Ala	Leu	Ala '		G1 y 505	Tyr	Ala	Gln .		G1n 510	Glu	Pro

5	As	р А1	a Ar 51	g G1 5	y Ar	g Il	e Ar	g As 52		y Al	a Le	u Le	u Ar 52		1 Ту	r Va
	Pr	o Ar 53	g Se 10	r Se	r Le	u Pr	o G1 53		е Ту	r Ar	g Th	r Se 54		u Th	r Le	u Al
10	A1 54	a Pr 5	o G1	u Ala	3 A1	a G1 55		u Va	1 G1	u Ar	g Lei 555		e Ġ1	y Hi	s Pr	o Lei 561
	Pr	o Le	u Ar	g Lei	Ası 569	o A1	a II	e Th	r Gl	y Pro 570		; G1	u G1	v G1	y G1 57	
15	Lei	- G1	u Thi	r Ile 580	. Leu	G1;	y Tr	p Pro	585			ı Arç	y Th	r Va 590		l Ile
	Pro	Se	r Ala 595	ı Ile	Pro	Thi	- Asp	600		j Asn	Va1	G1 y	G1)		Lec	ı Asp
20	Pro	Se:	- Ser	Ile	Pro	Asp	615		G1n	Ala	Ile	Ser 620		Leu	Pro	Asp
25	Tyr 625	Αla	Ser	Gln	Pro	G1 y 630		Pro	Pro	Arg	G1 u 635	Asp	Pro	Leu	Ala	Pro 640
	G1 y	Ile	Pro	Ser	G1 y 645	Pro	Leu	Lys	Ala	G1 u 650	Ile	Ala	Gln	Arg	Lev 655	
30	Asp	.Va1	Phe	Ala 660	G1 y	Lys	Asn	Thr	Asp 665	Leu	G1u	Val	Lev	Met 670	Glu	Trp
	Leu	Lys	Thr 675	Arg	Pro	Пe	Leu	Ser 680	Pro	Leu	Thr	Lys	G1 y 685	Ile	Leu	Gly
35	Phe	Va1 690	Phe	Thr	Leu	Thr	Va1 695	Pro	Ser	Glu		G1 y 700	Leu	G1n	Arg	Arg
40	Arg 705	Phe	Val	Gln	Asn	Ala 710	Leu	Asn	Gly		G1 y 715	Asp	Pro	Asn	Asn	Met 720
· =	Asp	Lys	Ala		L y s 725	Leu	Tyr	Arg		Leu 730	Lys	Arg	G1 u		G1 y 735	Lys
45	Pro	Pro		Glu # 740	Asp (	Leu	Lys		G1 u 745	Phe						

	(2) INFORMATION FOR SEQ ID NO:30:	
5	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 14 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
15	CTAGACTAGT CTAG	. 14
	(2) INFORMATION FOR SEQ ID NO:31:	
20	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 15 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:	
	GGCGGCAGAA AGAGC	15
	(2) INFORMATION FOR SEQ ID NO:32:	
35	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 21 amino acids</li> <li>(B) TYPE: amino acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
40	(ii) MOLECULE TYPE: peptide	
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
	Met Lys Ala Asn Leu Leu Val Leu Leu Cys Ala Leu Ala Ala Asp 1 5 10 15	
50		

	Ala Asp Thr Ile Cys 20	
5	(2) INFORMATION FOR SEQ ID NO:33:	
10	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 72 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	,	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	,
20	GGCAGAAAGA TGAAGGCAAA CCTACTGGTC CTGTTATGTG CACTTGCAGC TGCAGATGCA GACACAATAT GC	6 7
	(2) INFORMATION FOR SEQ ID NO:34:	
25	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 24 amino acids</li> <li>(B) TYPE: amino acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
30	(ii) MOLECULE TYPE: peptide	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:  Gly Arg Lys Met Lys Ala Asn Leu Leu Val Leu Cys Ala Leu Ala 1 5 10 15	
40	Ala Ala Asp Ala Asp Thr Ile Cys 20	
	(2) INFORMATION FOR SEQ ID NO:35:	
45	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 63 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	

(ii) MOLECULE TYPE: DNA (genomic)

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
	ATGAAGGCAA ACCTACTGGT CCTGTTATGT GCACTTGCAG CTGCAGATGC AGACACAATA	6
10	TGA	6:
	(2) INFORMATION FOR SEQ ID NO:36:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 21 amino acids  (B) TYPE: amino acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
20	(ii) MOLECULE TYPE: peptide	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:  Met Lys Ala Asn Leu Leu Val Leu Leu Cys Ala Leu Ala Ala Asp 1 5 10 15	
30	Ala Asp Thr Ile Xaa 20 (2) INFORMATION FOR SEQ ID NO:37:	
35	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 27 amino acids  (B) TYPE: amino acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
40	(ii) MOLECULE TYPE: peptide	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
45	His His Ala Asn Glu Asn Ile Phe Tyr Cys Pro Ile Ala Ile Met Ser 1 5 10 15	
	Ala Leu Ala Met Val Tyr Leu Gly Ala Lys Asp 20 25	
50	<b>25</b>	

50

	(2) INFORMATION FOR SEQ ID NO:38:	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 81 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
	CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG	60
	GTATACCIGG GTGCAAAAAG C	81
20	(2) INFORMATION FOR SEQ ID NO:39:	
25	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 27 amino acids</li> <li>(B) TYPE: amino acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
	(ii) MOLECULE TYPE: peptide	
30		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
35	His His Ala Asn Glu Asn Ile Phe Tyr Cys Pro Ile Ala Ile Met Ser l 5 10 15	
	Ala Leu Ala Met Val Tyr Leu Gly Ala Lys Ser 20 25	
40	(2) INFORMATION FOR SEQ ID NO:40:	
45	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 78 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li></ul>	
	(D) TOPOLOGY: linear  (ii) MOLECULE TYPE: DNA (genomic)	
50		

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:	
	GGCAGAAAGA TGAAGGCAAA CCTACTGGTC CTGTTATGTG CACTTGCAGC TGCAGATGCA	60
5	GACACAATAT GCATGATG	78
	(2) INFORMATION FOR SEQ ID NO:41:	
10	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 26 amino acids  (B) TYPE: amino acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15	(ii) MOLECULE TYPE: peptide	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
	Gly Arg Lys Met Lys Ala Asn Leu Leu Val Leu Leu Cys Ala Leu Ala 1 5 10 15	
25	Ala Ala Asp Ala Asp Thr Ile Cys Met Met 20 25	
25	(2) INFORMATION FOR SEQ ID NO:42:	
30	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 72 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
35	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
40	GGCATGAAGG CAAACCTACT GGTCCTGTTA TGTGCACTTG CAGCTGCAGA TGCAGACACA	60
	ATATGCATGA TG	72
45		

	(2) INFORMATION FOR SEQ ID NO:43:	
5	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 24 amino acids</li> <li>(B) TYPE: amino acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
10	(ii) MOLECULE TYPE: peptide	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
	Gly Met Lys Ala Asn Leu Leu Val Leu Cys Ala Leu Ala Ala Ala 1 5 10 15	
20	Asp Ala Asp Thr Ile Cys Met Met 20	
	(2) INFORMATION FOR SEQ ID NO:44:	
25	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 90 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
30	(ii) MOLECULE TYPE: DNA (genomic)	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:  GTATGCATGC ACCATGCCAA TGAGAACATC TTCTACTGCC CCATTGCCAT CATGTCAGCT	60
	CTAGCCATGG TATACCTGGG TGCAAAAGAC	90
40	(2) INFORMATION FOR SEQ ID NO:45:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 30 amino acids	
<b>4</b> 5	(B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

	(11) NOLECOLE 1172: peptide	
5	(wi) SEQUENCE DESCRIPTION, SEQ 10 Mg. 45	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
	Val Cys Met His His Ala Asn Glu Asn İle Phe Tyr Cys Pro Ile Ala 1 5 10 15	
10	Ile Met Ser Ala Leu Ala Met Val Tyr Leu Gly Ala Lys Asp 20 25 30	
	(2) INFORMATION FOR SEQ ID NO:46:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 147 base pairs  (B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
	ATGAAGGCAA ACCTACTGGT CCTGTTATGT GCACTTGCAG CTGCAGATGC AGACACAATA	.66
30	TGCCACCATG CCAATGAGAA CATCTTCTAC TGCCCCATTG CCATCATGTC AGCTCTAGCC	120
	ATGGTATAÇC TGGGTGCAAA AGACAGC	147
	(2) INFORMATION FOR SEQ ID NO:47:	
35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 49 amino acids	
	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single	
40	(D) TOPOLOGY: linear	
••	(ii) MOLECULE TYPE: peptide	
45		

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
5	Met Lys Ala Asn Leu Leu Val Leu Leu Cys Ala Leu Ala Ala Ala Asp 1 5 10 15	
	Ala Asp Thr Ile Cys His His Ala Asn Glu Asn Ile Phe Tyr Cys Pro 20 25 30	
10	Ile Ala Ile Het Ser Ala Leu Ala Het Val Tyr Leu Gly Ala Lys Asp 35 40 45	
	Ser	
15	(2) INFORMATION FOR SEQ ID NO:48:	
	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 70 base pairs</li><li>(B) TYPE: nucleic acid</li></ul>	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEO ID NO:48:	
	CCTATCAGAA ACGAATGGGG GTGCAGATGC AACGGTTCAA GCGCGAGGAC CTGAAGTAAG	60
30	AATTCGAGCT	70
	(2) INFORMATION FOR SEQ ID NO:49:	
35	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 2013 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
40	(ii) MOLECULE TYPE: DNA (genomic)	
45		

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

5	ATGGCCGAGG AAGCTTTCGA CCTCTGGAAC GAATGCGCCA AAGCCTGCGT GCTCGACCTC	60
	AAGGACGGCG TGCGTTCCAG CCGCATGAGC GTCGACCCGG CCATCGCCGA CACCAACGGC	120
	CAGGGCGTGC TGCACTACTC CATGGTCCTG GAGGGCGGCA ACGACGCGCT CAAGCTGGCC	180
10	ATCGACAACG CCCTCAGCAT CACCAGCGAC GGCCTGACCA TCCGCCTCGA AGGCGGCGTC	240
	GAGCCGAACA AGCCGGTGCG CTACAGCTAC ACGCGCCAGG CGCGCGGCAG TTGGTCGCTG	300
15	AACTGGCTGG TACCGATCGG CCACGAGAAG CCCTCGAACA TCAAGGTGTT CATCCACGAA	360
	CTGAACGCCG GCAACCAGCT CAGCCACATG TCGCCGATCT ACACCATCGA GATGGGCGAC	420
	GAGTTGCTGG CGAAGCTGGC GCGCGATGCC ACCTTCTTCG TCAGGGCGCA CGAGAGCAAC	480
20	GAGATGCAGC CGACGCTCGC CATCAGCCAT GCCGGGGTCA GCGTGGTCAT GGCCCAGACC	540
	CAGCCGCGCC GGGAAAAGCG CTGGAGCGAA TGGGCCAGCG GCAAGGTGTT GTGCCTGCTC	600
25	GACCCGCTGG ACGGGGTCTA CAACTACCTC GCCCAGCAAC GCTGCAACCT CGACGATACC	660
	TGGGAAGGCA AGATCTACCG GGTGCTCGCC GGCAACCCGG CGAAGCATGA CCTGGACATC	720
	AAACCCACGG TCATCAGTCA TCGCCTGCAC TTTCCCGAGG GCGGCAGCCT GGCCGCGCTG	780
30	ACCGCGCACC AGGCTTGCCA CCTGCCGCTG GAGACTTTCA CCCGTCATCG CCAGCCGCGC	840
	GGCTGGGAAC AACTGGAGCA GTGCGGCTAT CCGGTGCAGC GGCTGGTCGC CCTCTACCTG	900
35	GCGGCGCGGC TGTCGTGGAA CCAGGTCGAC CAGGTGATCC GCAACGCCCT GGCCAGCCCC	960
00	GGCAGCGGCG GCGACCTGGG CGAAGCGATC CGCGAGCAGC CGGAGCAGGC CCGTCTGGCC	1020
	CTGACCCTGG CCGCCGCCGA GAGCGAGCGC TTCGTCCGGC AGGGCACCGG CAACGACGAG	1080
40	GCCGGCGCGG CCAACGCCGA CGTGGTGAGC CTGACCTGCC CGGTCGCCGC CGGTGAATGC	1140
	GCGGGCCCGG CGGACAGCGG CGACGCCCTG CTGGAGCGCA ACTATCCCAC TGGCGCGGAG	1200
	TTCCTCGGCG ACGGCGGCGA CGTCAGCTTC AGCACCCGCG GCAGTCTTCT AACCGAGGTC	1260
45	GAAACGTACG TICTCTCTAT CATCCCGTCA GGCCCCCTCA AAGCCGAGAT CGCACAGAGA	1320
	CTTGAAGATG TCTTTGCAGG GAAGAACACC GATCTTGAGG TTCTCATGGA ATGGCTAAAG	1380

50

	ACAAGACCAA TCCTGTCACC TCTGACTAAG GGGATTITAG GATTTGTGTT CACGCTCACC	1440
5	GTGCCCAGTG AGCGAGGACT GCAGCGTAGA CGCTTTGTCC AAAATGCCCT TAATGGGAAC	1500
5	GGGGATCCAA ATAACATGGA CAAAGCAGTT AAACTGTATA GGAAGCTCAA GAGGGAGATA	1560
	ACATTCCATG GGGCCAAAGA AATCTCACTC AGTTATTCTG CTGGTGCACT TGCCAGTTGT	1620
10	ATGGGCCTCA TATACAACAG GATGGGGGCT GTGACCACTG AAGTGGCATT TGGCCTGGTA	1680
	TGTGCAACCT GTGAACAGAT TGCTGACTCC CAGCATCGGT CTCATAGGCA AATGGTGACA	1740
	ACAACCAACC CACTAATCAG ACATGAGAAC AGAATGGTTT TAGCCAGCAC TACAGCTAAG	1800.
15	GCTATGGAGC AAATGGCTGG ATCGAGTGAG CAAGCAGCAG AGGCCATGGA GGTTGCTAGT	1860
	CAGGCTAGGC AAATGGTGCA AGCGATGAGA ACCATTGGGA CTCATCCTAG CTCCAGTGCT	1920
20	GGTCTGAAAA ATGATCTTCT TGAAAATTTG CAGGCCTATC AGAAACGAAT GGGGGTGCAG	1980
	ATGCAACGGT TCAAGCGCGA GGACCTGAAG TAA	2013
	(2) INFORMATION FOR SEQ ID NO:50:	
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 671 amino acids	
	(B) TYPE: amino acid (C) STRANDEDNESS: single	
30	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: peptide	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
	Met Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys 1 5 10 15	i
40	Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Het Ser Val Asp	)
	20 25 30	
	Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Het 35 40 45	•
45	Val teu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala	
	50 . 55 60	
50		
5 <b>5</b>		

	65		r II	e Thi	r Se	r Ası 70	p G1	y Lei	u Thi	r Ile	e Arg 75	g Le	u G1	u G1	y G1	y Va1 80
5	Gl	u Pro	o Ası	n Ly:	85	va Va	l Ar	g Tyı	r Sei	7 Tyı 90	r Thi	r Ar	g G1	n Ala	a Ar 95	g Gly
10	Se	r Trj	o Sei	r Lec 100		1 Trp	Lei	u Val	105		e G1,	y Hi	s G1	U Ly: 110		o Ser
	Ası	n Ile	9 Lys		l Ph∈	: Ile	e His	120		Asn	n Ala	G1	y Asi 12!		n Le	u Ser
15	His	130		r Pro	lle	: Tyr	135		e Glu	Met	. G1 y	/ Asp 14(		J Lei	. Le	u Ala
	Lys 145		, Ala	Arg	Asp	150		Phe	Phe	Va1	Arg 155		His	; G1 c	Se i	160
20	Glu	Met	. G1n	Pro	Thr 165		Ala	Ile	Ser	His 170		G1y	/ Val	Ser	· Val	Val
25	Met	. Ala	Gln	Thr 180	Gln	Pro	Arg	Arg	61 u 185	Lys	Arg	Trp	Ser	· G1u 190		Ala
	Ser	Gly	Lys 195		Lev	Cys	Leu	Leu 200	Asp	Pro	Leu	Asp	G1 y 205		Tyr	Asn
30		Leu 210	Ala	Gln	Gln	Arg	Cys 215	Asn	Leu	Asp	Asp	Thr 220	Trp	G1'u	Gly	Lys
	11e 225		Arg	Val	Leu	A1a 230	G1 y	Asn	Pro	Ala	Lys 235	His	Asp	Leu	Asp	Ile 240
35	Lys	Pro	Thr	Val	I1e 245	Ser	His	Arg	Lev	His 250	Phe	Pro	GΊυ	G1 <sub>.</sub> y	G1 y 255	Ser
40	Leu	Ala	Ala	Leu 260	Thr	Ala	His	G1n	A1a 265	Cys	His	Leu	Pro	Leu 270	G1u	Thr
40	Phe	Thr	Arg 275	His	Arg	Gln	Pro	Arg 280	Gly	Trp	Glυ	G1n	Leu 285	Glu	G1n	Cys
45	G1 y	Tyr 290	Pro	Val	G1 n	Arg	Leu 295	Va1	Ala	Leu	Tyr	Leu 300	Ala	Ala	Arg	Leu
	Ser 305	Trp	Asn	GIn		Asp 310	G1n	Va 1	Ile		Asn 315	Ala	Leu	Ala	Ser	Pro 320

	Gly	Ser	G1 y	G1 y	Asp 325	Leu	G1 y	G1v	Ala	11e 330	Arg	Glu	Gln	Pro	G1 u 335	Gln
5	Ala	Arg	Lev	A1a 340	Lev	Thr	Lev	Ala	A1a 345	Ala	G1 u	Ser	G1 u	Arg 350	Phe	Val
	Arg	Gln	G1 y 355	Thr	G1 y	Asn	Asp	G1u 360	Ala	G] y	Ala	Ala	Asn 365	Ala	Asp	Val
10	۷a۱	Ser 370	Leu	Thr	Cys	Pro	Va1 375	Ala	Ala	Gly	G1 u	Cys 380	Ala	Gly	Pro	Ala
45	Asp 385	Ser	Gly	Asp	Ala	Leu 390	Leu	Glu	Arg	Asn	Tyr 395	Pro	Thr	G1 y	Ala	61 u 400
15	Phe	Leu	Gly	Asp	G1 y 405	Gly	Asp	Val	Ser	Phe 410	Se <u>r</u>	Thr	Arg	G1 y	Ser 415	Leu
20	Leu	Thr	Glu	Va1 420	Glu	Thr	Tyr	Val	Leu 425	Ser	Ile	Ile	Pro	Ser 430	G1 y	Pro
	Leu	Lys	A1a 435		Ile	Δla	Gln	Arg 440		Glu	Asp	Val	Phe 445	Ala	G1 y	Lys
25	Asn	Thr 450		Leu	G1 u	Val	Leu 455		G1u	Trp	Leu	Lys 460	Thr	Arg	Pro	Ile
	Leu 465		· Pro	Leu	Thr	Lys 470		Ile	. Leu	ı Gly	Phe 475	Val	Phe	Thr	Leu	Thr 480
30	Val	Pro	) Ser	· Glu	485		Leu	Glr	n Arg	490	Arg	Phe	· Val	G1 n	Asn 495	Ala
35	Lei	, Ası	n G1)	Asr 500		Asp	) Pro	Asr	505		. Asp	Lys	; Ala	Va1	Lys	Leu
	Ту	r Ar	g Lys 515		. Lys	. Arg	G G I u	520		r Phe	His	; G1 y	, A1a 525	Lys	: G1u	ılle
40	Se	r Le		r Tyi	r Sei	r Ala	53!		a Le	u Ala	s Sei	- Cys 540		G1.y	/ Lei	Ile
	Ту .54		n Ar	g Me	L G1:	y A1a 550		Th:	r Th	r Glu	۷a` 55!	1 A1a 5	a Phe	e G1 y	, Le	Val 560
45	Су	s Al	a Th	r Cy:	s G1 (		n Il	e Al	a As	p Sei 571		n Hi:	s Ar	g Sei	r Hi: 57!	s Arg

	Gln Met Val Thr Thr Asn Pro Leu Ile Arg His Glu Asn Arg Met 580 585 590	
5	Val Leu Ala Ser Thr Thr Ala Lys Ala Met Glu Gln Met Ala Gly Ser 595 600 605	
	Ser Glu Gln Ala Ala Glu Ala Met Glu Val Ala Ser Gln Ala Arg Gln 610 615 620	
10	Met Val Gln Ala Met Arg Thr Ile Gly Thr His Pro Ser Ser Ser Ala 625 630 635 640	
15	Gly Leu Lys Asn Asp Leu Leu Glu Asn Leu Gln Ala Tyr Gln Lys Arg 645 650 655	
	Met Gly Val Gln Met Gln Arg Phe Lys Arg Glu Asp Leu Lys Xaa 660 665 670	
20	(2) INFORMATION FOR SEQ ID NO:51:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 38 base pairs	
25	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30	(ii) MOLECULE TYPE: DNA (genomic)	
3.0	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:	
35		38
	(2) INFORMATION FOR SEQ ID NO:52:  (i) SEQUENCE CHARACTERISTICS:	
40	(A) LENGTH: 81 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
	(ii) MOLECULE TYPE: DNA (genomic)	
45		

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:	
5	ATAGAATICI TACTICAGGI CCTCGCGATI GTCGTACTCC TCTGCATTGI CTCCGAAGAA	60
	ATAAGATCCT TCATTACTCA T	81
	(2) INFORMATION FOR SEQ ID NO:53:	
10	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 2754 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> </ul>	
15	(D) TOPOLOGY: linear	•
	(ii) MOLECULE TYPE: DNA (genomic)	
		ē
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:	
	ATGGCCGAGG AAGCTTTCGA CCTCTGGAAC GAATGCGCCA AAGCCTGCGT GCTCGACCTC	60
25	AAGGACGGCG TGCGTTCCAG CCGCATGAGC GTCGACCCGG CCATCGCCGA CACCAACGGC	120
25	CAGGGCGTGC TGCACTACTC CATGGTCCTG GAGGGCGGCA ACGACGCGCT CAAGCTGGCC	180
	ATCGACAACG CCCTCAGCAT CACCAGCGAC GGCCTGACCA TCCGCCTCGA AGGCGGCGTC	240
30	GAGCCGAACA AGCCGGTGCG CTACAGCTAC ACGCGCCAGG CGCGCGGCAG TTGGTCGCTG	300
	AACTGGCTGG TACCGATCGG CCACGAGAAG CCCTCGAACA TCAAGGTGTT CATCCACGAA	360
	CTGAACGCCG GCAACCAGCT CAGCCACATG TCGCCGATCT ACACCATCGA GATGGGCGAC	420
35	GAGTIGCTGG CGAAGCTGGC GCGCGATGCC ACCTTCTTCG TCAGGGCGCA CGAGAGCAAC	480
	GAGATGCAGC CGACGCTCGC CATCAGCCAT GCCGGGGTCA GCGTGGTCAT GGCCCAGACC	540
40	CAGCCGCGCC GGGAAAAGCG CTGGAGCGAA TGGGCCAGCG GCAAGGTGTT GTGCCTGCTC	600
	GACCCGCTGG ACGGGGTCTA CAACTACCTC GCCCAGCAAC GCTGCAACCT CGACGATACC	660
	TGGGAAGGCA AGATCTACCG GGTGCTCGCC GGCAACCCGG CGAAGCATGA CCTGGACATC	720
45	AAACCCACGG TCATCAGTCA TCGCCTGCAC TTTCCCGAGG GCGGCAGCCT GGCCGCGCTG	780
	ACCGCGCACC AGGCTTGCCA CCTGCCGCTG GAGACTTTCA CCCGTCATCG CCAGCCGCGC	840
50		

GGCTGGGAA	L AALIGGAGLA	GIGCGGCIAI	CCGGIGCAGC	. GGC1GG1CGC	CCICIACCIG	900
GCGGCGCGG	C TGTCGTGGAA	CCAGGTCGAC	CAGGTGATCO	GCAACGCCCT	GGCCAGCCCC	960
GGCAGCGGC	G GCGACCTGGG	CGAAGCGATC	CGCGAGCAGC	CGGAGCAGGC	ссетстеесс	1020
CTGACCCTG	G CCGCCGCCGA	GAGCGAGCGC	TTCGTCCGGC	AGGGCACCGG	CAACGACGAG	1080
GCCGGCGCG	G CCAACGCCGA	CGTGGTGAGC	CTGACCTGCC	CGGTCGCCGC	CGGTGAATGC	1140
GCGGGCCCG	G CGGACAGCGG	CGACGCCCTG	CTGGAGCGCA	ACTATCCCAC	TGGCGCGGAG	1200
TTCCTCGGC	G ACGGCGGCGA	CGTCAGCTTC	AGCACCCGCG	GCATGGCGTC	CCAAGGCACC	1260
AAACGGTCT	T ACGAACAGAT	GGAGACTGAT	GGAGAACGCC	AGAATGCCAC	TGAAATCAGA	1320
GCATCCGTC	G GAAAAATGAT	TGGTGGAATT	GGACGATTCT	ACATCCAAAT	GTGCACAGAA	1380
CTTAAACTCA	GTGATTATGA	GGGACGGTTG	ATCCAAAACA	GCTTAACAAT	AGAGAGAATG	1440
GTGCTCTCTG	G CTITTGACGA	AAGGAGAAAT	AAATACCTGG	AAGAACATCC	CAGTGCGGGG	1500
AAGGATCCTA	AGAAAACTGG	AGGACCTATA	TACAGAAGAG	TAAACGGAAA	GTGGATGAGA	1560
GAACTCATCO	TTTATGACAA	AGAAGAAATA	AGGCGAATCT	GGCGCCAAGC	TAATAATGGT	1620
GACGATGCAA	CGGCTGGTCT	GACTCACATG	ATGATCTGGC	ATTCCAATTT	GAATGATGCA	1680
ACTTATCAGA	GGACAAGGGC	TCTTGTTCGC	ACCGGAATGG	ATCCCAGGAT	GTGCTCTCTG	1740
ATGCAAGGTT	CAACTCTCCC	TAGGAGGTCT	GGAGCCGCAG	GTGCTGCAGT	CAAAGGAGTT	1800
GGAACAATGG	TGATGGAATT	GGTCAGGATG	ATCAAACGTG	GGATCAATGA	TCGGAACTTC	1860
TGGAGGGGTG	AGAATGGACG	AAAACAAGA	ATTGCTTATG	AAAGAATGTG	CAACATTCTC	1920
AAAGGGAAAT	TTCAAACTGC	TGCACAAAA	GCAATGATGG	ATCAAGTGAG	AGAGAGCCGG	1980
GACCCAGGGA	ATGCTGAGTT	CGAAGATCTC	ACTTTTCTAG	CACGGTCTGC	ACTCATATTG	2040
AGAGGGTCGG	TTGCTCACAA	<b>GTCCTGCCTG</b>	CCTGCCTGTG	TGTATGGACC	TGCCGTAGCC	2100
AGTGGGTACG	ACTTTGAAAG	AGAGGGATAC	TCTCTAGTCG	GAATAGACCC	TTTCAGACTG	2160
CTTCAAAACA	GCCAAGTGTA	CAGCCTAATC	AGACCAAATG	AGAATCCAGC	ACACAAGAGT	2220
CAACTGGTGT	GGATGGCATG	CCATTCTGCC	GCATTTGAAG	ATCTAAGAGT	ATTGAGCTTC	2280

	ATCAAAGG	GA C	GAAG	GTGGT	CC	CAAG	AGGG	AAG	CTTT	CCA	CTAG	AGGA	GT T	CAAA	TTGC	T	2340
5	TCCAATGA	AA A	TATG	GAGA	: TA	TGGA	ATCA	AGT	ACAC	TTG	AACT	GAGA	AG C	AGGT	ACTG	G	2400
	GCCATAAG	GA C	CAGA	AGTG	G AG	GAAA	CACC	AAT	CAAC	AGA	GGGC.	ATCT	GC G	GGCC	AAAT	С	2460
	AGCATACA	AC C	TACG	TTCTO	AG	TACA	GAGA	AAT	CTCC	CTT	TTGA	CAGA	AC A	ACCG	TAT	G	2520
10	GCAGCATT	CA C	TGGG	AATA(	: AG	AGGG	GĄGA	ACA	TCTG	ACA	TGAG	GACC	GA A	ATCA	TAAG	G	2580
	ATGATGGA	AA G	TGCA	AGACO	AG	AAGA	TGTG	TCT	TTCC	AGG	GGCG	GGGA	GT C	TTCG	AGCT	С	2640
15	TCGGACGA	AA A	GGCA	GCGAC	cc.	CGAT	CGTG	ССТ	TCCT	TTG	ACAT	GAGT	AA T	GAAG	GATC	T	2700
	TATTTCTT	CG G	AGAC	AATGO	AG	AGGA	GTAC	GAC	AATC	GCG .	AGGA	ECTG	AA G	TAA			2754
	(2) INFO	RMAT:	101	FOR S	EQ	ID N	0:54	:				•					
20	(i)	•		E CHA					_								
		(B	) TYI	NGTH: PE: a	min	o ac	i d		5							•	
				RANDE POLOG			-	16							1		
25	(ii)	MOLE	ECUL	E TYF	'E: 1	prot	ein										
30	(xi)	SEQ	JENCE	E DES	CRI	PT IOI	N: S1	EQ II	ON C	:54:							
	Met 1	Ala	G1υ	G1υ	Ala 5	Phe	Asp	Leu	Trp	Asn 10	Glu	Cys	Ala	Lys	A1a 15	Cys	
		Lau	A = =	Leu		<b>A-n</b>	61	V-1	۸		Sar	Ara	Mat	Sar	_	A	
35	٧٥١	reu	ASP	20	Lys	MSh	оту	vai	25	361	Jei	Arg	Het	30	Vai	wsh	
	Pro	Ala	Ile 35	Ala	Asp	Thr	Asn	G1 y 40	G1n	G1 y	۷a۱	Leu	His 45	Tyr	Ser	Met	
40	V=3	Lau		G1 y	C1	4	A		1	t		A1-		A	A	A1-	
	٧٩١	50	910	Giy	u,	A311	55	MIA	Ceu	Lys	560	60	rie	wsh	M311	MIG	
	Leu 65	Ser	Ile	Thr	Ser	Asp 70	Gly	Leu	Thr	Ile	Arg 75	Lev	Glu	G1 y	Gly	Va1 80	
45		Dro	A.s.n	Lys	Dro		Ara	Tun	50=	Tur		A=a	Gl n	A1-	A = a		
	3.0	710		-	85	<b>va</b> 1	Ary	191	361	90	1111	ni y	3111	MIA	95	ч	
50																	

Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val 165  Met Ala Gln Thr Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala 180  Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn 200  Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys 210  Tile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile		Se	r Tr	p Sei	100		ı Trp	Lec	υ Va`	1 Pro		e G1	y Hi:	s G1:	u Ly:		) Ser
130 135 140  Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn 160  145 Glu Het Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val 165 175  Met Ala Gln Thr Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala 180 180 200  Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn 200  Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys 210 225  Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys 220 225  Lys Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser 220 225  Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 260 265  Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Ala Leu Arg Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Asn Ala Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gla Gly Gly Gly Ser 320  Ala Arg Leu Ala Asp Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val	5	Ası	n Il			Phe	: I1e	e His			ASI	n Ala	a G1			Leu	, Ser
Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn 160  Glu Het Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val 165  Het Ala Gln Thr Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala 180  Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn 200  Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys 215  Z10  Z21  Z25  Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile 225  Lys Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser 245  Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 260  Phe Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Glr Cys 290  Phe Thr Arg His Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu 290  Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln 335  Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val		His			· Pro	Ile	: Tyr			e 61 c	Me 1	: G1 <sub>3</sub>			J Lec	, Leu	ı Ala
15	10			u Ala	Arg	Asp			Phe	? Phe	· Val			His	G Tu	Ser	Asn 160
Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn 200   Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys 210   Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp 240   240   240   240   245   255	15	Glu	, Met	t G1n	Pro			Ala	ΙÌε	Ser			( G1 y	Val	Ser		
20		Met	. Ala	a Gln		Gln	Pro	Arg	Arg			Arg	Trp	Ser			Ala
210 215 220 216 216 217 218 220 221 221 222 223 223 223 223 225 226 2245 225 226 225 226 227 227 227 228 228 229 229 229 229 229 229 229 229	20	Ser	Gly		Val	Leu	Cys	Leu			Pro	Leu	Asp			Tyr	Asn
Lys Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser 245  Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 260  Phe Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys 275  Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu 290  Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Gln Gln Gln 335  Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val		Tyr			Gln	Gln	Arg		Asn	Leu	Asp	Asp			GΊυ	G1 y	Lys
Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 260 Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 270 Thr 27	25		Tyr	Arg	Val	Leu		Gly	Asn	Pro	Ala		His	Asp	Leu	Asp	Ile 240
Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 260 Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr 270 Thr 27	30	Lys	Pro	Thr	Val		Ser	His	Arg	Leu		Phe	Pro	G1u	G1 ý		Ser
Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu 290  Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln 325  Ala Arg Leu Ala Leu Thr Leu Ala Ala Glu Ser Glu Arg Phe Val	~	Leu	Ala	Ala		Thr	Ala	His	Gln		Cys	His	Leu	Pro		GΊυ	Thr
290 295 300  Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro 305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln 325  Ala Arg Leu Ala Leu Thr Leu Ala Ala Glu Ser Glu Arg Phe Val	35	Phe	Thr	Arg 275	His	Arg	G1n	Pro		G1 y	Trp	G1u	Gln		G1 y	Gln	Cys
305  Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln 325  Ala Arg Leu Ala Leu Thr Leu Ala Ala Glu Ser Glu Arg Phe Val		Gly		Pro	Va1	G1 n	Arg		Va1	Ala	Leu	Tyr		Ala	Ala	Arg	Lev
325 330 335  45  Ala Arg Leu Ala Leu Thr Leu Ala Ala Glu Ser Glu Arg Phe Val	40		Trp	Asn	G1 n			G1n	Val	Ile	Arg	Asn 315	Ala	Leu	Ala	Ser	
Ala Arg Leu Ala Leu Thr Leu Ala Ala Glu Ser Glu Arg Phe Val		Gly	Ser	G1 y			Leu	G1 y	Glu			Arg	Glu	Gln	Pro		G1 n
	45	Ala	Arg	Lev .	Ala ( 340	Lev	Thr	Leu			Ala	Glu	Ser			Phe	Val

	Arg	Gln	G1 y 355	Thr	Gly	Asn	Asp	G1u 360	Ala	Gly	Ala	Ala	Asn 365	Ala	Asp	Val
5	Val	Ser 370	Leu	Thr	Cys	Pro	Va1 375	Αla	Ala	Gly	G1υ	380 Çys	Ala	Gly	Pro	Ala
10	Asp 385	Ser	Gly	Asp	Ala	Leu 390	Leu	GΊυ	Arg	Asn	Tyr 395	Pro	Thr	Gly	Ala	G1 u 400
	Phe	Lev	Gly	Asp	G1 y 405	Gly	Asp	Val	Ser	Phe 410	Ser	Thr	Arg	Gly	Met 415	Ala
15	Ser	Gln	G1 y	Thr 420	Lys	Arg	Ser	Tyr	G1 u 425	Gln	Met	GΊυ	Thr	Asp 430	Gly	G1u `
	Arg	G1 n	Asn 435	ΑÌa	Thr	G1υ	Ile	Arg 440	Ala	Ser	Väi	Gly	Lys 445	Met	Ile	Gly
20	Gly	11e 450	G1 y	Arg	Phe	Tyr	11e 455	Gin	Met	Cys	Thr	G1u 460	Leu	Lys	Leu	Ser
	<b>Asr</b> 465	Tvr	èЯн	G1 y	Arg	Leu 470	Ile	Gln	Asn	Ser	Leu 475	Thr	Ile	Glu	Arg	Het 480
25	Val	Leu	Ser	Ala	Phe 485	Asp	G1u	Arg	Arg	Asn 490	Lys	Tyr	Leu	Glu	G1u 495	His
30	Pro	Ser	Ala	G1 y 500	Lys	Asp	Pro	Lys	Lys 505	Thr	G1 y	Gly	Pro	11e 510	Tyr	Arg
	Arg	Val	Asn 515	G1 y	Lys	Trp	Met	Arg 520	G1u	Leu	I1e	Leu	Tyr 525	Asp	Lys	Glυ
35	Glu	11e 530	Arg	Arą	Ile	Trp	Arg 535	G1n	Ala	Asn	Asn	G1 y 540	Asp	Asp	Ala	Thr
	A1a 545	G1 y	Lev	Thr	His	Me t 550	Met	Ile	Trp	His	Ser 555	Asn	Leu	Asn	Asp	A1a 560
40	Thr	Týr	Gln	Arg	Thr 565	Arg	Ala	Lev	Val	Arg 570	Thr	G1 y	Met	Asp	Pro 575	Arg
45	Het	Cys	Ser	Leu 580	Met	Gln	G1 y	Ser	Thr 585	Leu	Pro	Arg	Arg	Ser 590	Gly	Ala
	Ala	G1 y	A1a 595	Ala	Val	Lys	G1 y	Va1 600	G1 y	Thr	Met	Val	Me t 605	G1 u	Leu	Val
50																

	Arg	610		Lys	Arg	i-G1y	11e		Asp	Arg	Asn	620	•	Arg	, Gly	Glu
5	Asn 625		Arg	Lys	Thr	Arg 630		Ala	Tyr	G1 u	Arg 635		Cys	Asn	Ile	Leu 640
10	Lys	Gly	Lys	Phe	G1 n 645		Ala	Ala	G1n	Lys 650		Met	Met	Asp	G1n 655	
	Arg	G1 u	Ser	Arg 660		Pro	G1 y	Asn	A1a 665	Glu	Phe	Glu	Asp	Leu 670		Phe
15	Lev	Αīa	Arg 675	Ser	Ala	Leu	Ile	Leu 680		Gly	Ser	Val	Ala 685	His	Lys	Ser
	Cys	Leu 690		Ala	Cys	Val	Tyr 695	G1 y	Pro	Ala	Va1	700	Ser	G1 y	Tyr	Asp
20	Phe 705	G1 v	Arg	Glu	G1 y	Tyr 710	Ser	Leu	Val	G1 y	I1e 715	Asp	Pro	Phe	Arg	Leu 720
	Leu	G1 n	Asn	Ser	G1 n 725	Val	Tyr	Ser	Leu	11e 730	Arg	Pro	Asn	Glu	Asn 735	Pro
25	Ala	His	Lys	Ser 740	G1n	Leu	۷a٦	Trp	Met 745	Ala	Cys	His	Ser	A1 a 750	Ala	Phe
30	Glu	Asp	Leu 755	Arg	Val	Leu	Ser	Phe 760	Ile	Lys	Gly	Thr	Lys 765	Va1	Va1	Pro
		G1 y 770	Lys	Leu	Ser	Thr	Arg 775	Gly	Va1	Gln	Ile	A1a 780	Ser	Asn	Glυ	Asn
35	Met 785	G1 u	Thr	Met	G1u	Ser 790	Ser	Thr	Lev	Glu	Leu 795	Arg	Ser	Arg	Tyr	Trp 800
	Ala	Ile	Arg	Thr	Arg 805	Ser	G1 y	G1 y	Asn	Thr 810	Asn	Gln	Gln	Arg	A1a 815	Ser
40	Ala	G1 y		Ile 820	Ser	Ile	G1n	Pro	Thr 825	Phe	Ser	Val		Arg 830	Asn	Leu
45	Pro	Phe	Asp 835	Arg	Thr	Thr		Met 840	Ala	Ala	Phe		G1 y 845	Asn	Thr	Glu
	G1y .	Arg 850	Thr	Ser	Asp		Arg 855	Thr	G1 u	Ile		Arg 860	Met	Met	Glu	Ser
50																

	Ala Arg Pro Glu Asp Val Ser Phe Gln Gly Arg Gly Val Phe Glu Leu 865 870 875 880	
5	Ser Asp Glu Lys Ala Ala Ser Pro Ile Val Pro Ser Phe Asp Met Ser 885 890 895	
10	Asn Glu Gly Ser Tyr Phe Phe Gly Asp Asn Ala Glu Glu Tyr Asp Asn 900 905 910	
,,	Arg Glu Asp Leu Lys Xaa 915	
15	(2) INFORMATION FOR SEQ ID NO:55:	
13	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 35 base pairs</li><li>(B) TYPE: nucleic acid</li></ul>	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:	
	ATACCCGCGG CATGGGTGCG AGAGCGTCGG TATAT	35
30	(2) INFORMATION FOR SEQ ID NO:56:  (i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 36 base pairs	
35	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:	
	ATAGAATTCT CATTGTGACG AGGGGTCGCT GCCAAA	36
45		
50		

## (2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERIST	ICS:
---------------------------	------

(A) LENGTH: 2814 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

# (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

15	ATGAAAAAGA CAGCTATCGC GATTGCAGTG GCACTGGCTG GTTTCGCTAC CGTAGCGCAG	60
	GCCGCGAATT TGGCCGAAGA AGCTTTCGAC CTCTGGAACG AATGCGCCAA AGCCTGCGTG	120
20	CTCGACCTCA AGGACGGCGT GCGTTCCAGC CGCATGAGCG TCGACCCGGC CATCGCCGAC	180
	ACCAACGGCC AGGGCGTGCT GCACTACTCC ATGGTCCTGG AGGGCGGCAA CGACGCGCTC	240
	AAGCTGGCCA TCGACAACGC CCTCAGCATC ACCAGCGACG GCCTGACCAT CCGCCTCGAA	300
25	GGCGGCGTCG AGCCGAACAA GCCGGTGCGC TACAGCTACA CGCGCCAGGC GCGCGGCAGT	360
	TGGTCGCTGA ACTGGCTGGT ACCGATCGGC CACGAGAAGC CCTCGAACAT CAAGGTGTTC	420
30	ATCCACGAAC TGAACGCCGG CAACCAGCTC AGCCACATGT CGCCGATCTA CACCATCGAG	480
30	ATGGGCGACG AGTTGCTGGC GAAGCTGGCG CGCGATGCCA CCTTCTTCGT CAGGGCGCAC	540
	GAGAGCAACG AGATGCAGCC GACGCTCGCC ATCAGCCATG CCGGGGTCAG CGTGGTCATG	600
35	GCCCAGACCC AGCCGCGCG GGAAAAGCGC TGGAGCGAAT GGGCCAGCGG CAAGGTGTTG	660
	TGCCTGCTCG ACCCGCTGGA CGGGGTCTAC AACTACCTCG CCCAGCAACG CTGCAACCTC	720
	GACGATACCT GGGAAGGCAA GATCTACCGG GTGCTCGCCG GCAACCCGGC GAAGCATGAC	780
<b>40</b>	CTGGACATCA AACCCACGGT CATCAGTCAT CGCCTGCACT TTCCCGAGGG CGGCAGCCTG	840
	GCCGCGCTGA CCGCGCACCA GGCTTGCCAC CTGCCGCTGG AGACTTTCAC CCGTCATCGC	900
45	CAGCCGCGCG GCTGGGAACA ACTGGAGCAG TGCGGCTATC CGGTGCAGCG GCTGGTCGCC	960
	CTCTACCTGG CGGCGCGGCT GTCGTGGAAC CAGGTCGACC AGGTGATCCG CAACGCCCTG	1020

50

5

10

	GCCAGCCCCG	GCAGCGGCGG	CGACCTGGGC	GAAGCGATCC	GCGAGCAGCC	GGAGCAGGCC	1080
	CGTCTGGCCC	TGACCCTGGC	CGCCGCCGAG	AGCGAGCGCT	TCGTCCGGCA	GGGCACCGGC	1140
5	AACGACGAGG	cceececeec	CAACGCCGAC	GTGGTGAGCC	TGACCTGCCC	GGTCGCCGCC	1200
	GGTGAATGCG	ceeccceec	GGACAGCGGC	GACGCCCTGC	TGGAGCGCAA	CTATCCCACT	1260
10	GGCGCGGAGT	TCCTCGGCGA	CGGCGGCGAC	GTCAGCTTCA	GCACCCGCGG	CATGGGTGCG	1320
	AGAGCGTCGG	TATTAAGCGG	GGGAGAATTA	GATAAATGGG	AAAAAATTCG	GTTAAGGCCA	1380
	GGGGGAAAGA	AACAATATAA	ACTAAAACAT	ATAGTATGGG	CAAGCAGGGA	GCTAGAACGA	1440
15	TTCGCAGTTA	ATCCTGGUCT	TTTAGAGACA	TCAGAAGGCT	GTAGACAAAT	ACTGGGACA6	1500
	CTACAACCAT	CCCTTCAGAC	AGGATCAGAA	GAACTTAGAT	CATTATATAA	TACAATAGCA	1560
20	GTCCTCTATT	GTGTGCATCA	AAGGATAGAT	GTAAAAGACA	CCAAGGAAGC	CTTAGATAAG .	1620
20	ATAGAGGAAG	AGCAAAACAA	AAGTAAGAAA	AAGGCACAGC	AAGCAGCAGC	TGACACAGGA	1680
	AACAACAGCC	AGGTCAGCCA	AAATTACCCT	ATAGTGCAGA	ACCTCCAGGG	GCAAATGGTA	1740
25	CATCAGGCCA	TATCACCTAG	AACTTTAAAT	GCATGGGTAA	AAGTAGTAGA	AGAGAAGGCT	1800
	TTCAGCCCAG	AAGTAATACC	CATGTTTTCA	GCATTATCAG	AAGGAGCCAC	CCCACAAGAT	1860
	TTAAATACCA	TGCTAAACAC	AGTGGGGGGA	CATCAAGCAG	CCATGCAAAT	GTTAAAAGAG	1920
30	ACCATCAATG	AGGAAGCTGC	AGAATGGGAT	AGATTGCATC	CAGTGCATGC	AGGGCCTATT	1980
	GCACCAGGCC	AGATGAGAGA	ACCAAGGGGA	AGTGACATAG	CAGGAACTAC	TAGTACCETT	2040
35	CAGGAACAAA	TAGGATGGAT	GACACATAAT	CCACCTATCC	CAGTAGGAGA	AATCTATAAA	2100
	AGATGGATAA	TCCTGGGATT	AAATAAAATA	GTAAGAATGT	ATAGCCCTAC	CAGCATTCTG	2160
	GACATAAGAC	AAGGACCAAA	GGAACCCTTT	AGAGACTATG	TAGACCGATT	CTATAAAACT	2220
40	CTAAGAGCCG	AGCAAGCTTC	ACAAGAGGTA	AAAAATTGGA	TGACAGAAAC	CTTGTTGGTC	2280
	CAAAATGCGA	ACCCAGATTG	TAAGACTATT	TTAAAAGCAT	TGGGACCAGG	AGCGACACTA	2340
45	GAAGAAATGA	TGACAGCATG	TCAGGGAGTG	GGGGGACCCG	GCCATAAAGC	AAGAGTTTTG	2400
	GCTGAAGCAA	TGAGCCAAGT	ΑΔΓΑΔΑΤΓΓΑ	GCTACCATAA	TGATACAGAA	AGGCAATITT	2460

	AGGAACCA	AA GAA	AAGACTG	T TA	AGTG	TTTC	AAT	TGTG	GCA A	AAGA	AGGG	CA C	ATAG	CCAA	A	2520
5	AATTGCAG	GG CC	CCTAGGA	A AA	AGGG	CTGT	TGG	AAAT	GTG (	GAAA	GGAA	GG A	CACC	AAAT	G	2580
	AAAGATTG	TA ·CT(	GAGAGAC	A GG	CTAA	TTT	TTA	GGGA	AGA	TCTG	GCCT.	TC C	CACA	AGGG	A	2640
	AGGCCAGG	GA AT	ттсттс	A GA	GCAG	ACCA	GAG	CCAA	CAG	cccc	ACCA	GA A	GAGA	GCTT	С	2700
10	AGGTTTGG	GG AA	GAGACAA	C AA	CTCC	стст	CAG	AAGC	AGG /	AGCC	GATA	GA C	AAGG.	AACT	G	2760
	TATCCTTT	AG CT	TCCCTCA	G AT	CACT	CTTT	GGC	AGCG	ACC (	CCTC	GTCA	CA A	TGA			2814
15	(2) INFO	RMATIO	ON FOR	SEQ	ID NO	0:58	:									
	(i)	•	ENCE CH													
		(B)	TYPE:	amin	o ac	id		>								
20			STRAND TOPOLO			-	le									
	(ii)	MOLEC	CULE TY	PE: I	prote	ein										
25																
	(xi)	SEQUE	ENCE DE	SCRII	PT 101	۷: S	EQ II	ON C	: 58 :							
	Met 1	Lys l	ys Thr	Ala 5	Ile	Ala	Пe	Ala	Va1	Ala	Leu	Ala	•	Phe 15	Ala.	
30	•	V-1 4	VI. CI.	-	A1-	<b>A</b>	1	41-	•	C1	41-	Dha			T	
	inr	Vai F	11a G1n 20	Ala	АТА		FEO	25	310	910	міа	rile	30 ASP	reu	111	
35	Asn		lys Ala 35	Lys	Αla	Cys	Va1 40	Leu	Asp	Leu	Lys	Asp 45	G1 y	۷a۱	Arg	
	£ a			c	V-1			41-	71.	41-			<b>A</b>	C1	C1 -	
	261	50	Arg Met	ser	vai	55	PFO	Ała	116	AIA	60 60	INF	ASI	ч	GIN	
40	G1 y 65	۷a۱ (	.eu His	Tyr	Ser 70	Met	Val	Leu	G1 u	G1 y 75	G1 y	Asn	Asp	Ala	Leu 80	
						4.3		_				•	٥,			
	Lys	Leu A	Ala Ile	Asp 85	Asn	Ala	Leu	Ser	90	inr	Ser	ASP	ыу	95	inr	
45	Ile	Arg L	eu Glu. 100	G1 y	G1 ý	Va1	Glu	_	Asn	Lys	Pro	۷a۱		Tyr	Ser	
	-	<b>.</b>		4.1	4.	67	•	105	•		<b>A</b> .	<b>.</b>	110	14. •		
50	lyr		rg Gln	Ala	Arg	ыу	Ser 120	ırp	5er	Leu	ASN	17p	Leu	val	Pro	

	Ile	G1 y 130	His	Glu	Lys	Pro	Ser 135		Ile	Lys	Val	Phe 140		His	Glu	Leu
5	Asn 145	Ala	Gly	Asn	Gln	Leu 150	Ser	His	Het	Ser	Pro 155		Tyr	Thr	Ile	G1 u 160
	Met	Gly	Asp	Glu	Leu 165	Leu	Ala	Lys	Lev	A1a 170	Arg	Asp	Äla	Thr	Phe 175	Phe
10	Val-	Arg	Ala	His 180	Glu	Ser	Asn	GΊυ	Met 185	Gln	Pro	Thr	Leu	Ala 190	Ile	Ser
15	His	Ala	61y 195	Val	Ser	Val	Va1	Met 200	Ala	Gln	Thr	Gln	Pro 205	Arg	Arg	Glu
	Lys	Arg 210	Trp	Ser	Glυ	Trp	Ala 215	Ser	Gly	Lyš	°Va1	Leu 220	Cys	Leu	Leu	Asp
20	Pro 225	Leu	Asp	Gly	Va1	Tyr 230	Asn	Tyr	Leu	Ala	G1n 235	Gln	Arg	Cys	Asn	Leu 240
	Asp	Asp	Thr	Ţŗp	G1u 245	G1 y	Lys	Ile	Tyr	Arg 250	Val	Leu	Ala	G1 y	Asn 255	Pro
25	Ala	Lys	His	Asp 260	Leu	Asp	Пe	Lys	Pro 265	Thr	Val	Ile	Ser	Hi s 270	Arg	Leu
30	His	Phe	Pro 275	Glυ	Gly	Gly	Ser	Leu 280	Ala	Ala	Leu	Thr	A1a 285	His	G1n	Ala
·	Cys	His 290	Leu	Pro	Ĺeu	G1 u	Thr 295	Phe	Thr	Arg	His	Arg 300	G1n	Pro	Arg	G1 y
35	Trp 305	Glu	Gln	Lev	G1u	G1n 310	Cys	G1 y	Tyr	Pro	Va1 315	Gln	Arg	Leu	Val	A1a 320
	Leu	Tyr <sub>.</sub>	Leu	Ala	A1 a 325	Arg	Leu	Ser	Trp	Asn 330	G1n	Val	Asp	G1n	Va 1 335	Ile
40	Arg	Asn		Leu 340	Ala	Ser	Pro		Ser 345	G1 y	G1 y	Asp	Leu	GT y 350	G1 u	Ala
45	Ile	Arg	G1 u 355	Gln	Pro	Glu	G1n	A1a 360	Arg	Leu	Ala	Leu	Thr 365	Leu	Ala	Ala
	Ala	G1 u 370	Ser	Glu	Arg		Va1 375	Arg	G1n	G1 y	Thr	G1 y 380	Asn	Asp	G1u	Ala
50																

·	G1 y 385	Ala	Ala	Asn	Ala	Asp 390	Val	Val	Ser	Lev	Thr 395	Cys	Pro	Val	Ala	A1a 400
5	G1 y	GΊυ	Cys	Ala	G1 y 405	Pro	Ala	Asp	Ser	G1 y 410	Asp	Ala	Leu	Leu	G1 u 415	Arg
10	Asn	Tyr	Pro	Thr 420	Gly	Ala	GΊυ	Phe	Leu 425	Gly	Asp	G1 y	G1 y	Asp 430	Val	Ser
	Phe	Ser	Thr 435	Arg	Gly	Het	Gly	Ala 440	Arg	Ala	Ser	Val	Leu 445	Ser	G1 y	G1 y
15	Glu	Leu 450	Asp	Lys	Trp	Glu	Lys 455		Arg	Leu	Arg	Pro 460	G1 y	Gly	Lys	Lys
	G1n 465	Tyr	Lys	Leu	Lys	His 470	Ile	Val	Тгр	Ala	Ser 475	Arg	Glu	Leu	Glu	Arg 480
20	Phe	Ala	Val	Asn	Pro 485	G1 y	Leu	Lev	Glu	Thr 490	Ser	Glu	Gly	Cys	Arg 495	Gln
25	Ile	Leu	Gly	61 n 500	Lev	Gln	Pro	·Ser	Leu 505	Gln	Thr	G1 y	Ser	G1 u 510	Glu	Leu
	Arg	Ser	Leu 515	Tyr	Asn	Thr	Ile	A1a 520	Val	Leu	Tyr	Cys	Va1 525	His	G1n	Arg
30	Ile	Asp 530	Val	Lys	Asp	Thr	Lys 535	Glu	Ala	Leu	Asp	Lys 540	Ile	G1 u	Glu	Glu
	G1n, 545	Asn,	Lys	Ser	Lys	Lys 550	Lys	Ala	G1n	Gln	A1a 555	Ala	Ala	Asp	Thr	G1 y 560
35	Asn	Asn	Ser	G1 n	Va1 565	Ser	G1n	Asn	Tyr	Pro 570	Ile	Val	Gln	Asn	Leu 575	Gin
40				580					585					Asn 590		·
	Val	Lys	Va1 595	Va1	G1 u	Glu	Lys	A1a 600	Phe	Ser	Pro	G1 u	Va1 605	Ile	Pro	Met
45	Phe	Ser 610	Ala	Leu	Ser	Glu	G1 y 615	Ala	Thr	Pro	Gln	Asp 620	Leu	Asn	Thr	Met
	Leu 625	Asn	Thr	Val	G1 y	G1 y 630	His	G1n	Ala	Ala	Met 635	Gln	Met	Leu	Lys	G1 u 640
50																

## EP 0 541 335 A1

	Thṛ	Ile	Asn	Glυ	G1 u 645	Ala	Ala	Glu	Trp	Asp 650	Arg	Leu	His	Pro	Va1 655	His
5	Ala	Gly	Pro	11e 660	Ala	Pro	G1 y	G1n	Het 665	Arg	Glu	Pro	Arg	G1 y 670	Ser	Asp
10	Ile	Ala	G1 y 675	Thr	Thr	Ser	Thr	Leu 680	Gln	Glu	Gln	Ile	G1 y 685	Trp	Met	Thr
	His	Asn 690	Pro	Pro	Ile	Pro	Va1 695	G1 y	G1υ	Ile	Tyr	Lys 700	Arg	Trp	Ile	Ile
15	Leu 705	Gly	Leu	Asn	Lys	Ile 710	Val	Arg	Met	Tyr	Ser 715	Pro	Thr	Ser		Leu 720
	Asp	Ile	Arg	Gln	G1 y 725	Pro	Lys	Glu	Pro	Phe 730	Arg	Asp	Tyr	Val	Asp 735	Arg
20	Phe	Tyr	Lys	Thr 740	Leu	Arg	6fA	Glu	G1n 745	Ala	Ser	G1n	Glυ	Va1 750	Lys	Asn
25	Trp	Met	Thr 755	Glυ	Thr	Leu	Leu	Val 760	G1n	Asn	Ala	Asn	Pro 765	Asp	Cys	Lys
	Thr	Ile 770	Leu	Lys	Αla	Leu	G1 y 775	Pro	Gly	Ala	Thr	Leu 780	Glu	Glu	Met	Met
30	Thr 785	Ala	Cys	Gln	Gly	Va1 790	G1 y	G1 y	Pro	Gly	His 795	Lys	Ala	Arg	Val	Leu 800
	Ala	Glυ	Ala	Met	Ser 805	Gln	Val	Thr	Asn	Pro 810	Ala	Thr	Ile	Met	Ile 815	G1 n
35	Lys	G1 y	Asn	Phe 820	Arg	Asn	Gln	Arg	Lys 825	Thr	Val	Lys	Cys	Phe 830	Asn	Cys
	G1 y		G1 u 835	G1 y	His	Ile	Ala	Lys 840	Asn	Cys	Arg	Ala	Pró 845	Arg	Lys	Lys
40		Cys 850	Trp	Lys	Cys		Lys 855		Gly			Met 860	-	Asp	Cys	Thr
45	G1 u 865	Arg	Gln	Ala	Asn	Phe 870	Leu	G1 y	Lys		Trp 875	Pro	Ser	His	Lys	G1 y 880
	Arg	Pro	G1 y		Phe 885	Leu	Gln	Ser	Arg	Pro 890	G1 u	Pro	Thr	Ala	Pro 895	Pro
50																

### EP 0 541 335 A1

Glu Glu Ser Phe Arg Phe Gly Glu Glu Thr Thr Thr Pro Ser Gln Lys 900 905 910

Gln Glu Pro Ile Asp Lys Glu Leu Tyr Pro Leu Ala Ser Leu Arg Ser 920 925

Leu Phe Gly Ser Asp Pro Ser Ser Gln Xaa 930

10

5

### Claims

15

25

30

40

45

50

A recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a
modified <u>Pseudomonas</u> exotoxin and a polypeptide that is exogenous to an antigen-presenting cell, said
hybrid capable of being at least partially presented on an antigen-presenting cell surface.

2. A recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide of viral origin, said hybrid capable of being at least partially presented on an antigen-presenting cell surface.

- 3. A recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide of viral origin, said hybrid being capable of being internalized by an antigen-presenting cell and further capable of being at least partially presented on the surface of Said antigen-presenting cell.
  - 4. A recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide of viral origin, said hybrid capable of being internalized by an antigen-presenting cell and further capable of being processed for at least partial presentation on the surface of said antigen-presenting cell, sufficiently to elicit an immune response by cytotoxic T lymphocytes.
- 5. A transformant harboring a recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide that is exogenous to an antigen-presenting cell, said hybrid capable of eliciting an immune response by cytotoxic T lymphocytes.
  - 6. A transformant harboring a recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide that is exogenous to an antigen-presenting cell said hybrid capable of being at least partially presented on an antigen-presenting cell surface.
  - 7. A transformant harboring a recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <u>Pseudomonas</u> exotoxin and a polypeptide of viral origin, said hybrid capable of being at least partially presented on an antigen-presenting cell surface.
  - 8. A transformant harboring a recombinant DNA segment comprising a nucleotide sequence coding for a hybrid protein comprising a modified <a href="Pseudomonas">Pseudomonas</a> exotoxin and a polypeptide of viral origin, said hybrid capable of being internalized by an antigen-presenting cell, and further capable of being at least partially presented on the surface of said antigen-presenting cell.
  - The recombinant DNA segment as claimed in any one of claims 1 to 4, wherein said modified <u>Pseudomonas</u> exotoxin lacks a functioning ADP ribosylating domain.
- 10. The recombinant DNA segment as claimed in claim 2, wherein said polypeptide of viral origin is a viral protein fragment comprising the matrix protein of influenza A virus.
  - 11. The recombinant DNA segment as claimed in claim 10, wherein said viral protein fragment comprises re-

### EP 0 541 335 A1

sidues 57 to 68 of the matrix protein of influenza A virus.

5

10

25

30

35

40

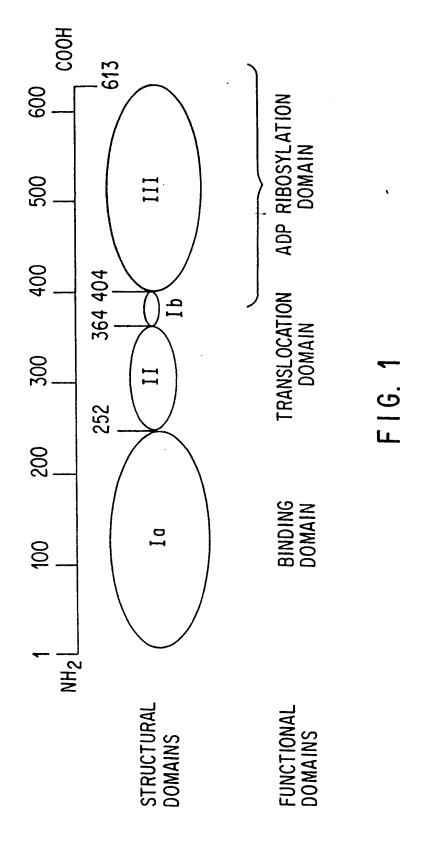
45

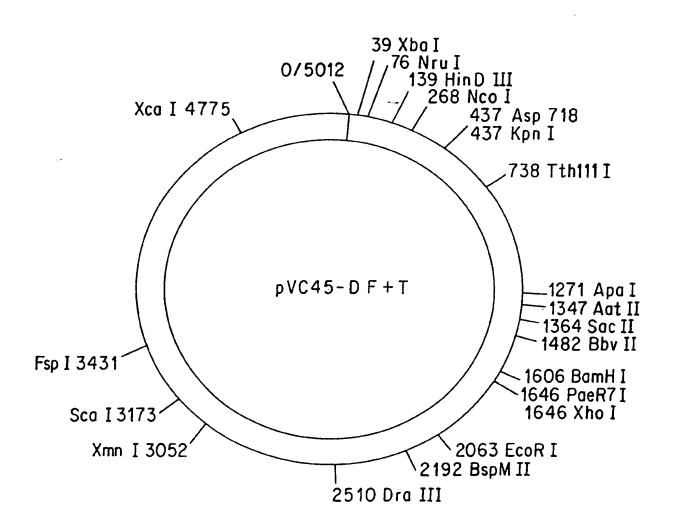
50

55

- 12. The recombinant DNA segment as claimed in claim 2, wherein said polypeptide of viral origin is a viral protein fragment comprising the gag protein of human immunodeficiency virus-1.
- 13. The recombinant DNA segment as claimed in claim 2, wherein said polypeptide of viral origin is a viral protein fragment comprising the nucleoprotein of influenza A virus.
- 14. The transformant as claimed in claim 5 wherein said modified <u>Pseudomonas</u> exotoxin lacks a functioning ADP ribosylating domain.
- 15. The transformant as claimed in claim 7, wherein said polypeptide of viral origin is a viral protein fragment comprising the viral matrix protein of influenza A virus.
- 16. The transformant as claimed in claim 15, wherein said viral protein fragment comprises residues 57 to 68 of the matrix protein of influenza A virus.
  - 17. The transformant as claimed in claim 7, wherein said polypeptide of viral origin is a viral protein fragment which is sufficiently specific to bind to HLA-2.
- 20 18. The transformant as claimed in claim 7, wherein said polypeptide of viral origin is a viral protein fragment comprising the nucleoprotein of influenza A virus.
  - 19. The transformant as claimed in claim 7, wherein said polypeptide of viral origin is a viral protein fragment comprising the gag protein of human immunodeficiency virus-1.

75





F1G. 2

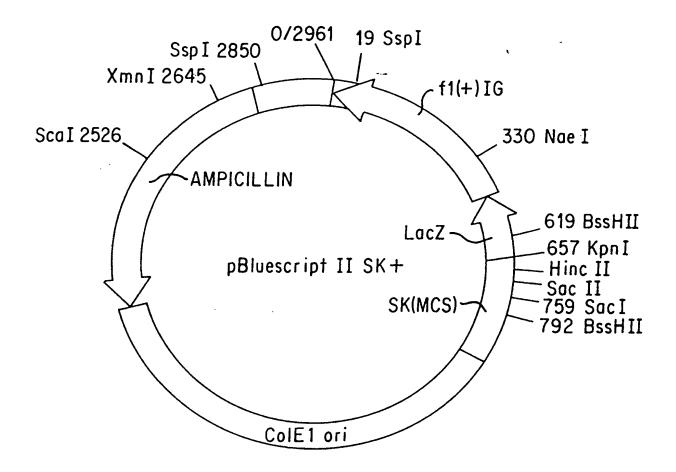


FIG. 3

THE DACE LEFT RI ANIV

٠.

ď

. ·

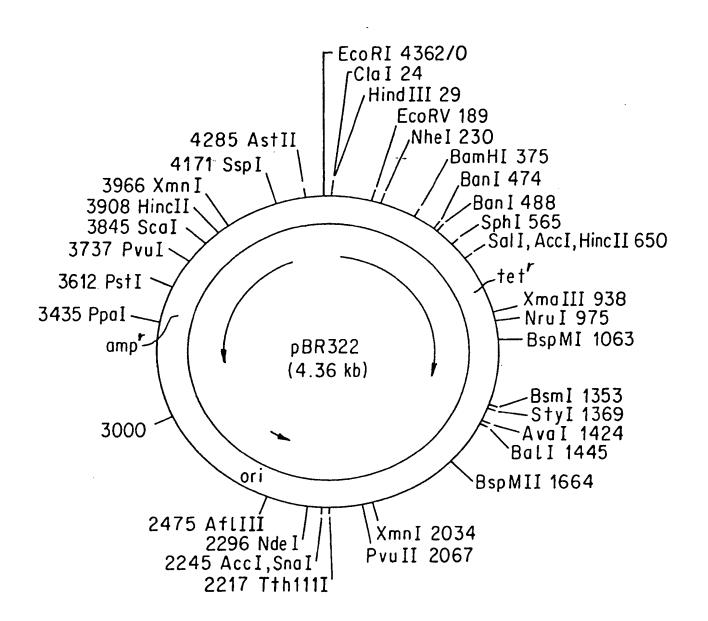
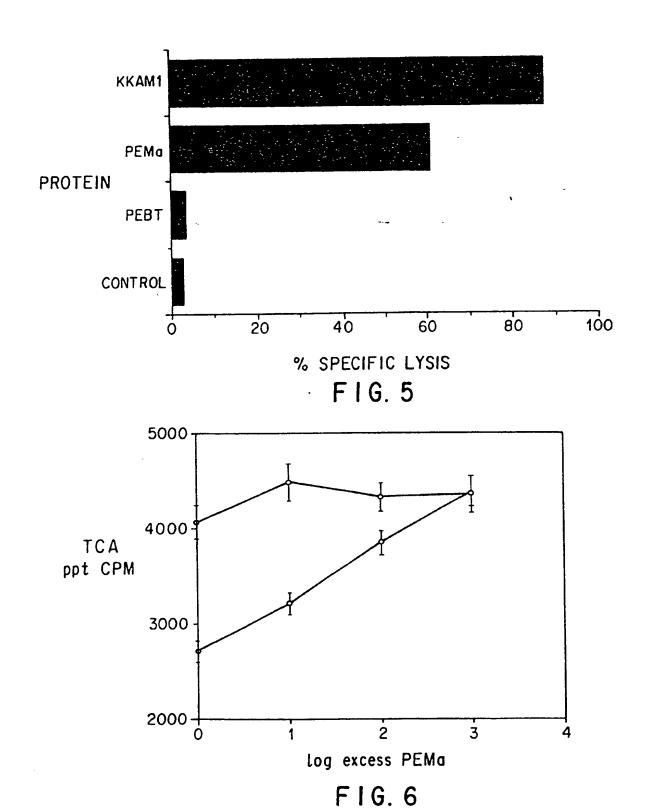


FIG. 4





# **EUROPEAN SEARCH REPORT**

Application Number

D	OCUMENTS CONSI	EP 92310067.1			
ategory	Citation of document with in of relevant pas	dication, where appropriate,	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. CL.5)	
A	EP - A - 0 261 (I. PASTAN et * Claims 1	al.)	1-4,9-	C 12 N 15/62 C 12 N 5/10	
A	EP - A - 0 431 (HOECHST AG) * Claims 1	· <del></del> -	1-19		
A ·	no. 7, Februar Columbus, Ohio T. ZEHAVI-WILI of murine cyto lymphocytes by aeruginosa exc page 564, colu abstract-no. 9 & Infect 56(1), 21	D, USA LNER "Induction olytic T Pseudomonas otoxin A", Imn 2, 64 194h Immun. 1988, 13-18	1-19	TECHNICAL FIELDS SEARCHED (Int. CL5)  C 12 N 5/00 C 12 N 15/00 A 61 K 37/00 A 61 K 39/00 C 07 K 15/00 C 12 P 21/00	
	The present search report has	Date of completion of the	: search	Examiner	
	VIENNA	19-02-1993		SCHARF	
X : part Y : part doc: A : tech O : non	CATEGORY OF CITED DOCUMI icularly relevant if taken alone icularly relevant if combined with a unent of the same category inological background -written disclosure rmediate document	E : earlie after D : document L : document	ber of the same patent fat	ioa ns .	